# A COMPUTATIONAL FRAMEWORK FOR INDIAN SIGN LANGUAGE RECOGNITION

*Thesis Submitted to*

*Cochin University of Science and Technology*

*For the award of the degree of*

***Doctor of Philosophy***

*Under*

***Faculty of Technology***

*By*


**Daleesha M Viswanathan**


*Under the Supervision of*

**Dr. Sumam Mary Idicula**


**DEPARTMENT OF COMPUTER SCIENCE**

**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**Kochi – 682022**


***August 2015***

# A COMPUTATIONAL FRAMEWORK FOR INDIAN SIGN LANGUAGE RECOGNITION

*Author*

***Daleesha M Viswanathan***
*Department of Computer Science,*
*Cochin University of Science and Technology*
*daleesha@cusat.ac.in*

*Supervisor*

***Dr. Sumam Mary Idicula***
*Professor and Head*
*Department of Computer Science,*
*Cochin University of Science and Technology*
*sumam@cusat.ac.in*

*August 2015*

**DEPARTMENT OF COMPUTER SCIENCE**
**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**
**COCHIN-682022, KERALA, INDIA**

# *Certificate*

*This is to certify that the thesis entitled* ***"A Computational Framework for Indian Sign Language Recognition"*** *is a bonafide record of the research carried out by* ***Daleesha M Viswanathan*** *under my supervision and guidance at the Department of Computer Science, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy under the Faculty of Technology, Cochin University of Science and Technology.*

*Kochi*
*August, 2015*

***Dr. Sumam Mary Idicula***
*Supervising Guide*
*Department of Computer Science*
*Cochin University of Science and Technology,*
*Kochi-682022, Kerala*

**DEPARTMENT OF COMPUTER SCIENCE**
**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**
**COCHIN-682022, KERALA, INDIA**


# *Certificate*

*This is to certify that all the relevant corrections and modifications suggested by the audience during the pre-synopsis seminar and recommended by the Doctoral Committee of the candidate have been incorporated in the thesis entitled*

***"A Computational Framework for Indian Sign Language Recognition"***

*Kochi*
*August, 2015*


***Dr. Sumam Mary Idicula***
*Supervising Guide*
*Department of Computer Science*
*Cochin University of Science and Technology,*
*Kochi-682022, Kerala*

# Declaration

*I, **Daleesha M Viswanathan**, hereby declare that the thesis titled **"A Computational Framework for Indian Sign Language Recognition"**, submitted to Cochin University of Science and Technology under Faculty of Technology is the outcome of the original research done by me under the supervision and guidance of Dr. Sumam Mary Idicula, Professor and Head, Department of Computer Science, Cochin University of Science and Technology. I also declare that this work did not form part of any dissertation submitted for the award of any degree, diploma, associateship, or any other title or recognition from any University or Institution.*

***Daleesha M Viswanathan***

# *Acknowledgement*

*I would like to thank Cochin University of science and Technology for the opportunity to pursue my PhD research in an ideal learning environment.*

*I express my sincere gratitude from the bottom of my heart to my guide Dr. Sumam Mary Idicula, Professor and Head, Department of Computer Science, Cochin University of Science and Technology, for the dedicated supervision and support during the entire span of this research.*

*I am thankful to Sr Abaya, St. Clara Oral Higher Secondary School for the Deaf, Manickamangalam for introducing me to the world of Indian Sign Language and the practical issues associated with it.*

*A special thanks to the faculty members, administrative staff and technical staff for the support rendered during this period.*

*I would like to thanks my fellow colleagues in Division of Information Technology and particularly Binsu C Kovoor for always kept me motivated during this entire journey.*

*A special appreciation is extended to all children who perform signs for this research purpose.*

*My deepest gratitude to my husband Sudheesh Manalil and my son Arjun for their support and inspiration during the entire span of my study. I also remember and thank my parents and sister for their love and support for me.*

*My salutations to God for empowering me to do a sincere work on a socially relevant topic.*

***Daleesha M Viswanathan***

## *Abstract*

Sign language is the primary means of communication for the hard to hear and speak people around the globe. Sign language emphasizes on visual possibilities as the participants are unable to hear sound patterns. Sign language uses different signs, body postures and gestures as opposed to sound patterns for communication, and evolves like any other spoken language. American Sign Language (ASL), British sign language (BSL), Arabic sign language (ArSL), Chinese sign language (CSL) and Indian sign language (ISL) are some of the widely used sign language systems around the world. There exists significant variation between sign languages, and due to these inherent variations, it is not possible to fully adopt a methodology that is found suitable for all. There are enormous complexities in ISL. Contrary to ASL, ISL sentences follow Subject-Object-Verb pattern. For example, the relative positioning of hand on face with respect to nose can convey 'WOMAN' or 'THINK' in ISL. Such complexities necessitate independent research in ISL.

Sign language recognition involves integration of different categories of signs. The signs can be mainly categorized into three groups like static hand gestures, dynamic gestures and facial expression. This research focuses on these three different channels and work to identify the potential of different computational methods to address some of the associated complexities with each channel. These complexities include static gestures with resemblances, static overlaid gestures, differential movement and directional changes in dynamic gestures and facial expression changes.

This research work is specifically focused to find a robust feature extraction method for sign representation. The existing feature extraction methods were compared for their potential to handle complexities like gesture resemblances, overlaid gestures etc. Different feature descriptors were tested for gestures with resemblances to identify the best descriptor. Semi global approach was followed instead of holistic or local approach.

This study also addressed the complexity in overlaid gestures. The overlaid gestures have two complexities - resemblances in gestures and complexity due to textural characteristics. This study tried a combination of feature descriptors rather than using them alone. The combination of HOG and LBP descriptors was found promising in addressing these complexities. In addition, a novel approach was tested in dynamic gesture recognition by using orientation features. In addition, this work also explored the recognition of facial expression changes in ISL sentences, which is a research area where not much work has been done.

This research tried, the recognition of simple ISL sentences following Subject-Object-Verb pattern. The literature survey and the analysis revealed the research gaps in ISL. The current research opens some of the future research directions that would improve ISL recognition. The evolved methods and combinations of them can be applied on a broad database of gestures so that precision and accuracy of gesture recognition can be improved substantially.

# Contents

# *List of Figures*

# *List of Tables*

# *List of Symbols*

| | |
|---|---|
| G | Gradient magnitude |
| $\theta$ | Orientation |
| $I_x$ | Horizontal gradients of the image *I* |
| $I_y$ | Vertical gradients of the image *I* |
| $LBP_{P,R}$ | LBP value for the central pixel of the cell where P is the total number of neighbours involves and R is the radius of the neighbourhood |
| $f(\theta_1)$ | Feature vector corresponding to orientation angle θ1 |
| $f(\theta_2)$ | Feature vector corresponding to directional angle θ2 |
| $\psi(x,y,f,\theta)$ | Gabor filter |
| $A_{(u,v)}(x,y)$ | Magnitude response |
| $\sigma$ | Standard deviation |
| $G_{(u,v)}(x,y)$ | Complex convolution output which decomposed into real and imaginary part |

## *List of Abbreviations*

| | |
|---|---|
| ASL | American Sign Language |
| BSL | British Sign Language |
| CSL | Chinese Sign Language |
| ArSL | Arabic Sign Language |
| Auslan | Australian Sign Language |
| ISL | Indian Sign language |
| PSL | Persian Sign language |
| ANN | Artificial Neural Networks |
| HOG | Histograms of Orientation Gradients |
| LBP | Local Binary Pattern |
| K-NN | K Nearest Neighbor classifier |
| SVM | Support Vector Machine |
| RBF | Radial Basis Function |
| PCA | Principal Component Analysis |
| SIFT | Scale Invariant Feature Transform |
| SURF | Speeded Up Robust Feature |
| HMM | Hidden Markov Model |

# Chapter 1

# INTRODUCTION

Contents



## 1.1 Overview

Sign language is the primary means of communication for the hard to hear and speak people around the globe. Sign language put emphasis on visual possibilities as the participants are unable to hear sound patterns. Sign language uses different signs, body postures and gestures as opposed to sound patterns for communication, and evolved like any other spoken language. Around 5% of people around the world are with hearing and speaking impairments and sign languages help these people to communicate and help to be part of the society by engaging in various activities. According to World Health Organization reports (WHO 2001), around 255 million people globally suffer from significant auditory loss, of which India accounts around three million people with varying levels of auditory loss. This has social implications because of the educational disparity and economic disadvantages suffered by these people, and eventually a substantial portion of the human resource gets marginalized. It is inevitable to have a sign language platform that is simple, efficient and accessible to these marginalized people so that they can be successful not only at personal level but also at societal level. Therefore, global, attempts have been made to integrate technological advancement and sign language linguistic advancements to provide better communication aids to the speaking and hearing impaired people.

Sign language communication involves not only hand gestures (manual signs) but also non_ manual signs conveyed through facial expressions, head movements and body postures. For example, in order to understand a signed sentence as shown in Fig 1.1: 'Are You Studying?', the signals in both, the manual and non-manual channels (facial expression) need to be recognized and fused.


**Figure 1.1** Sign Representation

Around 177 sign language systems are available globally. Few are American Sign Language (ASL), British Sign Language (BSL), Chinese Sign Language (CSL), Australian Sign Language (Auslan) and so on. Sign language evolved naturally like any other spoken language and always not have a strong connection with the native language. For example, British Sign Language (BSL) and American Sign Language (ASL) differ significantly although both the countries are native English speaking nations. As an example, ASL is mainly single handed while

BSL uses both hands. In addition, ASL has a strong connection with French sign language and Arabic sign language. However, ASL varies significantly with BSL and Australian sign language. Each sign language has its own independent standardization with in that country. Sign languages are well structured languages with phonology, morphology, syntax and grammar. Initial breakthrough in sign language research was in ASL which has remained as one of the most investigated sign languages in terms of linguistic structure. However, total adoption of methods of one sign language system to another system would not be always possible due to the inherent complexities of each. The main problem is that sign languages differ from country to country and region to region and signs and gestures are not uniform across different language systems. Sign Languages are:

- NOT the same all over the world
- NOT just 'languages of hands' alone. It also contains non-manual gestures.
- NOT just gestures, but also do have their own grammar.

## 1.2 Indian Sign Language

As indicated earlier, sign language may not have a strong connection with the local prevailing language. This is highly relevant in Indian context because of the multitude of regional languages and the regional variations of signs and syntax. Many efforts were made to codify the signs prevalent in India. In 1977, a linguistic analysis was performed by Vasishta, Woodword and Wilson by collecting signs from four major urban centres (Delhi, Calcutta, Bombay and Bangalore), which resulted in four dictionaries of Indian Sign Language (ISL) with regional variations. Ramakrishna Mission Vidyalaya, Coimbatore (2009) and IIT Guahatti (2012) lead an attempt to standardise ISL. In that study, Ramakrishna Mission Vidyalaya gathered 2037 signs from diversified sources (42 cities in 12 states) to provide a common sign language code for all over India indicating the complexity of sign language database. In addition, the standardization and research efforts are important in ISL primarily because not much research efforts have been carried out in ISL, although more than three million people suffer from auditory problems and more than ten million suffer from speaking difficulties. There are many intrinsic complexities in ISL primary due to its own syntax, morphology and grammar. In ISL both static and dynamic gestures are used for signing and the relative positioning of hand on face or body conveys complex meaning in ISL. Although ASL is single handed for alphabets, both hands are used in ISL for alphabets and other signs. Most importantly, subject–object-verb pattern is followed in ISL, contrary to subject- verb-object pattern in ASL. India has four distinct linguistic families of which ISL mostly follow Dravidian Language pattern.

The accessibility of Indians to ISL is low due to many social, economic and technological reasons. Less research attempts in ISL linguistics and technological advancement have resulted in the non-availability of simple ISL learning tools. Scarcity of sign language interpreter is also another reason for the backwardness of ISL literacy in India. Therefore, it is utmost important to empower the hard to hear and speak communities in India through technological intervention and refinement and development of appropriate research tools that aid these people in effective communication and learning. In addition, research efforts and refining of methods are required to correctly identify and judge proper interaction systems.

## 1.3 Challenges in Indian Sign Language

Following serious challenges are faced by hearing and speaking impaired people while trying to enter into educational, social and work environment.

**i. Languages and Available resources***:*

There are 22 officially recognized languages in India, of which 76% are originated or classified as Indo-Aryan language. ISL is based on Dravidian family of languages. However regional attempts are made by researchers to develop frameworks based on regional languages. So standardization of ISL is of utmost importance. Indeed, significant efforts are being made by Ali Yavar Jung National Institute for the Hearing Handicapped (AYJNIHH, Mumbai), Ramakrishna Mission Vidyalaya, Coimbatore (2009) and IIT Guahatti (2012) in this direction. These efforts should continue to have a common data set of signs for ISL.

### ii. Refining of proven methods in an Indian Context:

ISL is highly complicated and diverse in terms of signs and gestures used. The existing methods might vary their performance while handling complex gestures. So considering the enormous complexities of ISL, it is important to refine methods for better performance. Therefore different methods are needed to be screened for their potential to handle gestures with resemblances, dynamic gestures and non-manual gestures.

### iii. Modelling of Signs:

Shape and orientation of hand, hand motion and relative position of it with respect to body parts are the basic elements for sign language recognition. In addition, facial expressions and body postures would also supplement in conveying meaningful ideas. Hence integration of all the signs is an important challenge in the design of an ISL framework.

**iv. Environmental Conditions and Feature Extraction:**

The environmental conditions would significantly affect the performance of an image recognition framework. The illumination, shade due to overlapping images and angle of illumination would attribute experimental errors. Most of the methods are compared and tested under ideal uniform environment and environment interaction needs to be accounted while judging a method as superior or inferior.

## 1.4 Main Research Question

Due to the complexity and multimodal nature of ISL, the research area of ISL involves pattern recognition, machine learning, computer vision, natural language processing and linguistics. Thus, the main research question is:

> ***How to develop a framework that enables the recognition, processing and integration of the different channels of ISL communication?***

**Sub Research Questions**

Tackling this challenging problem raises a number of important sub-research questions like:

- *How to make computer understand different gestures?*
- *What representations or features are most appropriate for gesture identification, especially for complex gestures?*
- *How much efficiency can be achieved by using machine learning techniques?*
- *Can the feature identified for gesture recognition be used for developing a framework for ISL sentence recognition, which is the combination of all classes of gestures?*

## 1.5 Objectives and Scope

The objectives of this research work are identified as follows:

- *Identification of features that can represent simple and complex gestures using image processing techniques*

- *Develop a framework for Indian Sign Language sentence Recognition.*
- *Evaluate the performance of the framework.*

## 1.6 Contributions of the Thesis

The contribution of this thesis can be summarized as follows:

a) A gesture recognition scheme for ISL was designed and developed. The system included different phases like pre-processing, feature extraction, training and classification or recognition. Gestures with resemblance, overlaid gestures, dynamic gestures and facial expressions were considered for the study.

b) In the process of gesture recognition, the decisive feature descriptors are identified for each class of gestures with respect to the shape and style. The impact of these feature descriptors was studied through the experiments conducted on data sets.

c) A system for the recognition of simple ISL sentences was developed.

## 1.7 Outline of the Thesis

The thesis is organized in seven chapters as follows:

**Chapter 1** introduces ISL with its background and history. A brief discussion on the complexity, importance, challenges and research gaps in ISL along with the motivation, research questions and objectives of the research work are given in this chapter.

**Chapter 2** contains literature review of the works available till date in Sign Languages. The chapter outlines an overview of gesture recognition schemes available in American Sign Language, Arabic Sign Language, Chinese Sign language and ISL.

**Chapter 3** describes the salient features of static gestures in ISL. This chapter evaluated a set of feature descriptors and their combination that can be effectively used in the recognition of static gestures.

**Chapter 4** describes an integrated gesture recognition scheme for dynamic gestures.

**Chapter 5** describes an identification approach for facial expression changes in isolated ISL sentences.

**Chapter 6** describes the development of a simple ISL sentence recognition system. A grammar formalism was developed for simple ISL sentences. As a prior step for ISL sentence recognition, gesture spotting in compound words was done and it is discussed in this chapter. This chapter also explores and evaluates the recognition capability of the proposed framework.

**Chapter 7** includes the summary of the research work carried out, important contributions and details of possible future directions of work in this field.

*Chapter* **2**

# SIGN LANGUAGE RECOGNITION: A REVIEW

Contents



*Sign Languages have originated and evolved independently at different parts of the world. The various methods employed in sign language recognition on a global scale are reviewed in this chapter. Purpose of this survey is to examine data acquisition, feature extraction and classification methods employed for sign language recognition. Due to the interconnection of these areas, vast literature is available for review. Hence the domain for literature survey is restricted to sign language research – specifically Indian Sign language.*

## 2.1 Introduction

Advances in image processing and pattern recognition have influenced considerably the process of sign language recognition. It is important to have an automated sign language recognizer that can reduce the gap between common people and hard to hear, so that they can be brought into the main stream. Secondly, automated sign language recognizer would provide tutoring platform without a trained interpreter who are not widely available. Lastly, the knowledge gap between common people and hard to hear people will be minimum, and both can contribute to the development of the society.

Many works have been carried out in different sign languages around the world. Table 2.1 lists some of the well-known sign languages where active research works are being carried out. These sign languages have their own style or pattern of representation. So it is very clear that the problems associated with recognition differ across sign languages.

**Table 2.1:** Popular Sign languages used around the world.

| Sl.No | Country/Continent | Sign Language | Abbreviation |
|---|---|---|---|
| 1 | United States of America | American Sign language | ASL |
| 2 | United Kingdom | British Sign Language | BSL |
| 3 | Australia | Australian Sign Language | Auslan |
| 4 | Middle-East | Arabic Sign Language | ArSL |
| 5 | China | Chinese Sign language | CSL |
| 6 | Japan | Japanese Sign Language | JSL |
| 7 | Taiwan | Taiwanese Sign Language | TSL |

Among gesture categories, sign language is often regarded as the most structured one. Each Sign language in the world is a combination of manual and non-manual gestures with its own grammar. Category of gestures used in sign language is depicted in figure 2.1. In order to build a suitable automated sign language recognition system, a detailed interpretation of gestures is necessary. Due to the intrinsic differences existing across the sign languages, the approaches for their interpretation also differ.

Sign language recognition is not a simple task like speech recognition. The developments in sign language recognition are far behind speech recognition in terms of accuracy and correctness. Multiple channels are involved in sign language recognition contrary to audio channel recognition that is unidimensional and simple.

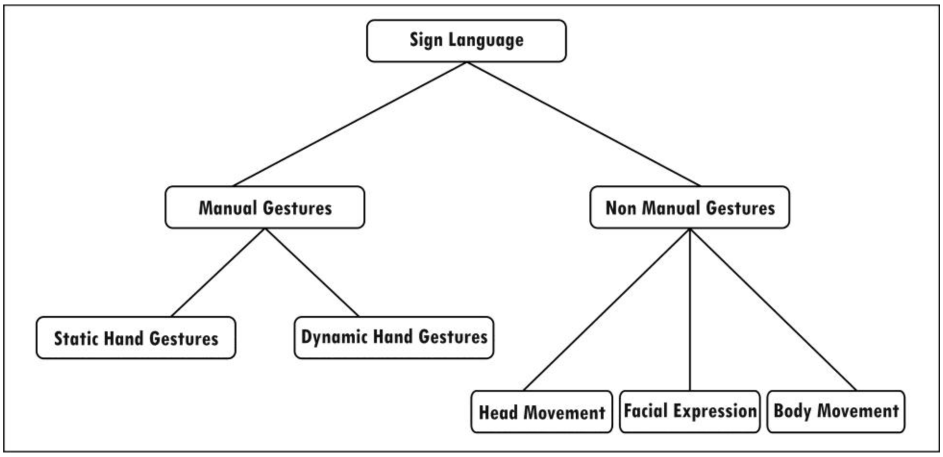



**Figure 2.1:** Category of gesture in sign language

The rest of the chapter is organized as follows. The state of art in the recognition of sign languages like ASL, ArSL and CSL is presented in detail in section 2.2. The performance evaluation of various gesture recognition schemes across multiple languages is discussed in 2.3. Furthermore, the progress made in ISL recognition in comparison to other languages is given in Section 2.4.

## 2.2 State of the Art in Sign Language Recognition

A survey on common methods using in gesture recognition is published in [1,2]. A detailed analysis on the recognition methods for various sign languages is reported by [3]. Figure 2.2 describes the general framework of a gesture recognition system.


**Figure 2.2:** General framework of a gesture recognition system.

Acquisition of image, feature extraction and its classification are the primary phases of gesture recognition. The raw image frames are pre-processed prior to analysis. Identifying the target image area is an important aspect of any recognition system. Identifying and extracting the most relevant features of a gesture plays an important role in increasing the accuracy of recognition systems. Vision based approach or direct measure approach is used in image acquisition. Data acquisition devices like Accelegloves are used in direct measure approach [4] or wearable computing approaches. However, vision based methods use skin colour or textural changes to track hand gestures [5]. Coloured gloves could also be used to track hands [6] (Fig 2.3).

**Figure 2.3:** Color glove and sensor glove

Sign language includes both manual and non-manual gestures as shown in Fig 2.1. All hand gestures are considered as manual gestures while facial expression, body movement and head movement are non-manual gestures. Since dynamic gestures follow a trajectory motion, the shape of hand and motion are to be considered. But static manual gestures are based on fixed hand postures. Sign language is a continuous gesture stream that is performed one after the other to ensure meaningful communication.

## 2.3 An Analysis of Methods Employed in Sign Language Recognition

Sign language communication is highly complex, and it has significant commonality with research in machine learning. Hand shape, location and motion trajectory of the hand and facial expressions are some of the main aspects relevant to machine analysis.

### 2.3.1 American Sign Language Recognition

Liang et al. [21] in 1995 made an initial attempt to develop a gesture recognition system for hearing impaired considering the wide use of ASL. The work focused on the recognition of a continuous flow of alphabets in ASL to spell a word. Sensor glove was used for capturing the gestures. Template matching recognition strategy was adopted for classification.

Gupta, Lalit et al [16] in 2001, have worked on a set of static hand gesture representing ASL signs. Features of the gesture were extracted based on the contour of the gesture. Classification was based on similarity measure.

Isaacs et al. [18] in 2004 focused on a gesture recognition system that utilized wavelet based feature vector for the recognition of 24 static ASL alphabets using still images. The pre-processing of images as well as mother wavelet for feature vector composition were optimized using Genetic Algorithms. Classification process was done using ANN classifier.

Oz, Cemil et al. [10] in 2007, presented an ASL word recognition system using Artificial Neural Networks (ANN) which converted ASL words to text. Sixty words were considered for this purpose. This system used sensory glove and 3D motion tracker to extract the features. Hand shape was defined based on joint angle between fingers, and the movement trajectory data was extracted by tracker. Signs were defined by hand shape, location, orientation, movement and distance. An accuracy rate of 95% was achieved by this approach.

Derpanis, Konstantinos G et al [26] in 2008 presented a paper to represent and recognize the hand movements that are used in single handed ASL. The approach followed was by decomposing the dynamic gestures into their static and dynamic components. Kinematic features were extracted from apparent motion used for identification for 14 primitive movements in ASL. This approach was evaluated on a database of 592 gesture sequences and yielded an overall recognition rate of 86%.

Ding, L et al. [12] in 2009, interpreted manual signs using hand shape, motion and place of articulation. Hand shape was represented as a set of affine equations. 3D motion path of the hand was tracked by the hand pose differences from the consecutive frames. The translation and rotation were estimated by using three point perspectives pose made by the object and the points were collected from camera coordinate system. Hand position was estimated by setting face as a reference. Classification was done using tree structure instead of HMM. The database included of 576 video sequences. Class of the sign was identified by integrating the three hand features.

Elmezain, Mahmoud et al. [13] in 2009, recognized ASL alphabet and Arabic numbers represented by single hand motion trajectory using HMM. System used combined feature of location, orientation and velocity. There were 720 video samples and 360 video sequences for training and for testing, respectively. Proposed system gained a recognition rate of 98.33%.

Alon et al. [7] in 2009 developed an ASL sign retrieval system to extract signs from a video sequence. They generated a framework for simultaneously performing spatial segmentation, temporal segmentation and recognition. The method was first applied to a hand motion in air to represent numbers. Hand motion tracking was done using frame difference method. Learning the observation density functions was done usinga variant of the Baum-Welch algorithm. This method achieved 85 % correct detection rate.

Athitsos et al. [2009] presented a database based approach for addressing ASL recognition [8]. Two gesture problem handled in this were hand shape and hand motion recognition. The hand motion has to be discriminated between the signs. Dynamic time wrapping distance measure was used in this analysis. Performance was evaluated using three measures: retrieval time, K-percentile accuracy and classification accuracy with 33 %.

Rashid et al. [23] in 2009 described an approach for hand posture recognition for static alphabets and numbers used in ASL. Segmentation of bare hand was exploited using normal Gaussian distribution information. Statistical and geometrical properties of the hand were treated as feature vectors. Hu moment invariant was considered for statistical feature vector generation. In order to avoid misclassifications in alphabets, curvature analysis was also carried out. SVM classifier was used for classification and recognition. Proposed framework gained an accuracy rate of 98.65% for ASL alphabet and 98.6% for numerals. The same authors tried another experiment using Microsoft Kinect sensor for capturing hand gestures [24]. Feature extraction was based on the depth and intensity of the image captured. Deep Belief Network was used for classification and recognition.

Kong et al. [17] in 2010 presented an approach to segment phonemes from ASL sentences. Hand motion trajectories of the signed sentences were segmented using rule base algorithm. Principal Component Analysis as feature descriptor was used to represent the segments. Training was done using Hidden Markov Model to recognize the sequence of the phonemes in the sentences. The average recognition error was 11.4%.

Ullahet et al. [14] in 2011 presented a research work based on Cartesian Genetic Programming (CGP) for learning ASL alphabet recognition system using. The average recognition accuracy was greater than 90%. .

Kim, Taehwan et al. [19] in 2012 presented a system for recognition of finger spelling sequences in ASL from video. Each signer has finger spelled words from a list of 300 words. Sixty image frames were taken from the video. System followed skin color based hand segmentation. Feature extraction of the segmented hand was generated using SIFT. PCA was applied on the feature vector for dimension reduction. Dimensionally reduced feature vector was taken by multilayer perception with one hidden layer having 1000 hidden nodes. Outputs of MLP were used as observations in HMM based recognizer.

Kurakin, Alexey et al. [20] in 2012 presented a paper on the recognition of dynamic hand gesture in ASL. Feature vector included of velocity of hand centre, rotation parameter of hand and shape descriptor. This study calculated the cell occupancy feature and silhouette feature from uniformly gridded hand image and applied PCA for feature dimension reduction. Training and testing was done using HMM.

Nguyen, Tan Dat et al. [25] in 2012 presented a paper for tracking and recognizing facial features exhibiting facial expressions used in ASL. This paper handled both head pose change and facial expression change in depicting a sign. Probabilistic Principal Component Analysis (PPCA) was used as shape vectors to learn the subspace transition probabilities for the tacking algorithm. Recognition framework was analyzed using nine HMMs and an SVM classifier and the study yielded an accuracy of 91.76%.

Bhat, Nagaraj N et al. [9] in 2013, proposed a method for static hand gesture recognition using radial enclosing of edge image and Self Organizing Map (SOM). Eighteen hand gestures were considered for this approach and this method had attained 92% recognition rate.

Tangsuksant et al [15] in 2014 used static hand postures representing ASL alphabets for sign recognition. This research designed a glove with six different color markers and developed algorithm for alphabet classification. The study used two camera to extract 3D coordinate points from each color marker to act as feature of the sign. Features were classified using feed forward Artificial Neural Network and yielded an accuracy of 95%.

Liu, Jingjing et al. [22] in 2014 proposed an automatic recognition system for non-manual grammatical markers based on head pose and facial expressions used in ASL. This paper analysed eyebrow raising and lowering, and different types of head movements such as head nods and shakes. Features are based on facial geometry and appearance along with head pose obtained through 3D deformable face tracker based on adaptive ensemble of Active Shape Models (ASM). Non manual event recognition was employed using two levels of CRF. Precision, recall and F1score values were more than 80%.

### 2.3.2 Arabic Sign Language Recognition

Shanableh et al. [40] in 2007 proposed feature extraction based on spatio-temporal feature of the ArSL gesture using 2-D Discrete Cosine Transform. HMM was used to classify images based on the temporal dependencies.

Shanableh et al. [31] in 2007 presented a variety of feature extraction methods for recognition of ArSL. Purpose of the system was to extract the sign representing images from the video stream and identify the extracted sign. The identified image was then transformed into the frequency domain and parameterized into a precise and concise feature sets. Classification was done using HMM and comparison was done using KNN and Bayesian classifiers.

Al-Rousan et al. [30] in 2009 introduced an ArSL recognition system based on HMM model. Thirty isolated words were used for this purpose. Feature extraction was done using Discrete Cosine Transform (DCT).This work achieved a recognition rate ranging from 90.6% to 98.13%.

Tolba, M. F et al. [32] in 2010 proposed a system to identify hand poses represented in a sentence of three words. The data set consisted of 30 sentences using 100 words. Feature extraction was done using pulse-coupled neural network (PCNN). Sign recognition was done using ‘‘graph-matching’’ algorithm. More than 70% recognition rate was achieved by this method.

Mohandes et al. [36] in 2013 presented a paper for two handed sign system using glove data. Features extracted from glove data and hands tracking based on decision level using Dempster Shafer theory were combined to represent the feature vector. The combined feature descriptor gained 98.1% accuracy rate for the recognition system.

Elons, A. S [33] in 2014 described a recognition system to identify six facial expressions used in ArSL. Feature extraction was done using Recursive Principle Components (RPCA). Multilayer Perceptron (MLP) was used for classification,. They also integrated facial expression with hand gestures and achieved 88% to 98% accuracy.

Mohandes et al. [34,35] in 2014 developed a system for Arabic alphabet sign recognition using Leap Motion Controller (LMC).Twenty eight Arabic alphabet signs with ten samples of each were collected from a single signer. Twelve features were extracted out of 23 values given by the LMC to represent each frame. Classification was done using Nave Bayes Classifier. NBC gave an accuracy rate of 98.3%.

Tubaiz et al. [29], in 2014 proposed a system having dataset of 40 sentences using 80 words. Two DG5-V hand data gloves were used to capture the hand gestures. Camera setup was used to synchronize hand gestures with their corresponding words. K-NN classifier was used for testing.

Al-Jarrah et al. [27] in 2015 proposed a recognition system for ArSL alphabets. Two feature extraction schemes namely boundary features and region features were computed and used for the representation of hand gesture. Boundary features were extracted as the length of line segments originating from the centroid of the hand gesture. Region features, were extracted after segmenting the hand gesture region into five clusters using k-means clustering technique. Adaptive Neuro-Fuzzy Inference System (ANFIS) model was used for training and testing and a recognition rate of 97.5 % was achieved on using 10 rules.

Tharwat, Alaa et al. [28] in 2015 used SIFT feature descriptors for representing static ArSL gestures. Linear Discriminant Analysis (LDA) was used for dimensionality reduction. Classifiers like SVM and K-NN were used for testing and this method achieved 99% recognition rate.

Aujeszky, Tamás et al. [37] in 2015 presented a paper using Microsoft Kinect device for gesture recognition process and attained 96% of accuracy. Aly, Saleh et.al. [38] in 2014 described dynamic gestures using spatiotemporal Local Binary Pattern feature vector. Data set consisted of 23 signs and classification was done using SVM classifier. This method gained 99.5% accuracy rate. Aly, Sherin et. al. [39] in 2014 analysed the same database using LBP with PCA for dimension reduction and training by HMM model.

### 2.3.3 Chinese Sign Language Recognition

Fang, G.L et al. [42, 43, 44, 45, 46] proposed different methods for Chinese Sign Language recognition system. Data glove based feature extraction and classification using self-organizing feature maps (SOFMs) with HMMon a database of 208 videos is described in [43]. They achieved 1.9% improvement in the accuracy rate compared to their previous work [42], which had an accuracy rate of 91.9%.

Quan, Y et al. [47] in 2010 extracted features based on temporal and spatial characteristics of a video sequence consisting of CSL manual alphabet images. Linear SVM classifier was used for identification process and the method achieved 99.7% recognition rate for letter 'F'.

Wang, C.L et al. [48] in 2002 extracted signs from sign data streams using Dynamic Programming (DP) and used ANN approach combining k-means for classification. Seventy one hand postures were used in this analysis. In [49], raw data were collected using Cyber Glove and a 3-D tracker. HMM was used for recognition purpose and an accuracy rate of over 90% was achieved.

Zhou, Y et al. [50] in 2008 used Volume Local Binary Patterns (VLBP) as feature descriptor and Polynomial Segment Model (PSM) to represent temporal evolution of sign features as a Gaussian process with time-varying parameter. This method outperformed conventional HMM methods by 6.81% in recognition rate. Zhou, Y et al. [51] in 2007 presented a method using etyma-based signer adaption for CSL vocabulary.

A summary of the review conducted is given in table 2.2 and 2.3.

**Table 2.2:** Manual Gestures Recognition Methods on Multiple Sign Languages

| Author | Sign Vocabulary | Features Extracted | Classification Methods | Accuracy | Language |
|---|---|---|---|---|---|
| Bhat, Nagaraj N., et al.[9] | ASL alphabets | Radial enclosing of egde of the image | Self-Organizing Map (SOM) | 92% | ASL |
| Karami,et al [63] | 32 static hand gestures PSL alphabets | Discrete Wavelet Transform (DWT) | Multi-Layer Perceptron (MLP) | 94% | Persian Sign Language (PSL) |
| Teng, Xiaolong, et al, [64] | 20 static hand gestures | Local Linear Embedding algorithm | Distance measure | 90% | CSL |
| Oz, Cemil, et al, [10] | 30 dynamic gestures | Sensor glove and 3D motion tracker | ANN | 95% | ASL |
| Derpanis, K.G.,et al [11] | Dynamic gestures with 14 motions | Kinematic features | KNN with Euclidian Distance measure | 86% | ASL |
| Ding, L.et a l[12] | Dynamic gestures | Translation and rotation by using three point perspective pose made by the hand | HMM | | ASL |
| Elmezain, Mahmoud, et al [13] | Motion trajectory | Features based in orientation, location and velocity | HMM | 98.33% | ASL |
| Rashid et al [23] | Static gestures | Hu moment invariant | SVM | 98.6% | ASL |
| Al-Rousan, M et al [30] | Static Isolated words | Discrete Cosine Transform | HMM | 90.6% to 98.13% | ArSL |
| Tolba, M. F.,et al [32] | Hand pose | Pulse coupled Neural Network | Graph matching algorithm | 70% | ArSL |
| Tharwat, Alaa, et al.[28] | Static gestures | Scale-Invariant Feature Transform | SVM | 99% | ArSL |
| Aujeszky, Tamás et al [37] | Dynamic Gestures, 23 signs | Local Binary Pattern | SVM | 99.5% | ArSL |
| Wang, C.L., et al [48] | Manual alphabets | Feature extraction based on Cyber glove | ANN & Dynamic programming by combining K-mean cluster | 90% | CSL |

**Table 2.3:**Non_ManualGesturesRecognition Methods on Multiple Sign Languages

| Author | Sign Vocabulary | Features Extracted | Classification Methods | Performance | Language |
|---|---|---|---|---|---|
| Liu, Jingjing, et al[22] | 1.Eyebrow raising and lowering 2. Head nods and shakes | 3D de face tracker based on an adaptive ensemble of formable of ASMs (Active Shape Model) | 2 level Conditional Random Field (CRF) | 80% | ASL |
| Nguyen, Tan Dat, et al [25] | Head pose change and facial expressional change in depicting sign | Probabilistic Principal Component Analysis (PPCA) | HMM and SVM | 91.76% | ASL |
| Elons, A. S., [33] | Six facial expressions used in ArSL | Recursive Principle Components(RPCA) | Multilayer Perceptron | Range between 88% to 98% | ArSL |
| Hrúz, M., J, [65] | Head move, facial expression,lip move | Active Shape Model (ASM) with land mark detector(LD) | HMM | 80% | General Sign Language |
| Von Agris,et al [66] | Facial expression and lip outline | Active Appearance Model | | 80.2% to 96.9% | German Sign Language |

## 2.4 Indian Sign language Recognition

So far very few studies in Indian Sign language have been documented. Table 2.4 illustrates the research work done in this area. Major works has been done for manual gestures. Especially in recognizing static gestures from still hand postures or spotting static gestures from continuous stream of gestures.

The graphical analysis presented in figure 2.4 and 2.5 clearly indicates the quantum of works carried out in ASL compared to other sign languages. Among the languages, lowest number of published research works was in ISL. The year wise analysis of research in ISL (Fig. 2.5) clearly indicates that proper emphasis for ISL was given only recently. Therefore, this emphazises the need for more research in this area.

**Table 2.4:** Indian Sign language Recognition

| Author, Year | Sign Vocabulary | Features Extracted | Classification Methods | Gesture Set | Performance |
|---|---|---|---|---|---|
| Nandy, Anup, et al, 2010 [52] | 22 words | Direction histogram with 18 bins and 36 bins | Euclidean distance and K-nearest neighbor metrics. | Manual, Static gestures | K-nearest neighbor gives good classification with 36 bin histogram with an accuracy ranging from 61.93% to 100% |
| Rekha, J.et al, 2011 [53] | Alphabets | Principle Curvature based region with wavelet packet decomposition | Hand postures are classified using Support Vector Machine. Dynamic gestures are classified using Dynamic Time Warping. | Manual, Static and dynamic gestures. | Static gestures with 91.3% and dynamic gestures with 86.3%. |
| Kishore, P,et al , 2011 [54] | Set of words | Elliptical Fourier descriptors for shape feature extraction and principal component analysis for feature set optimization and reduction. | Fuzzy classification | Manual, Static gestures | 91% |
| Tewari, et al, 2012, [55] | Static Alphabets | Two Dimension Discrete Cosine Transform (DCT) for each region is computed and feature vectors are formed from these DCT coefficients. | Kohonen Network | Manual, Static gestures | 80% |
| Singha et al,2013 [57] | Static Alphabets | Eigen Vectors | Euclidian distance | Manual, Static gestures | Average recognition rate of 97%. |
| Geetha, M.,et al, 2012[59] | Static Alphabet | Maximum Curvature Points as key frames for gesture shape identification | Support Vector Machine | Manual, Static gestures | Accuracy > 88% |
| Dour, Shweta et al, 2013 [58] | Static alphabets | Centre of gesture, distance of measure to boundary and degree measure as feature measures | Fuzzy classification | Manual, Static gestures. | Accuracy >68% |
| Geetha, M.,et al, 2013[61] | 10 words | Axis of Least Inertia is proposed for trajectory based feature extraction | Euclidian distance | Manual , Dynamic gestures | Compile time for each word is between 70 to 130 seconds. |

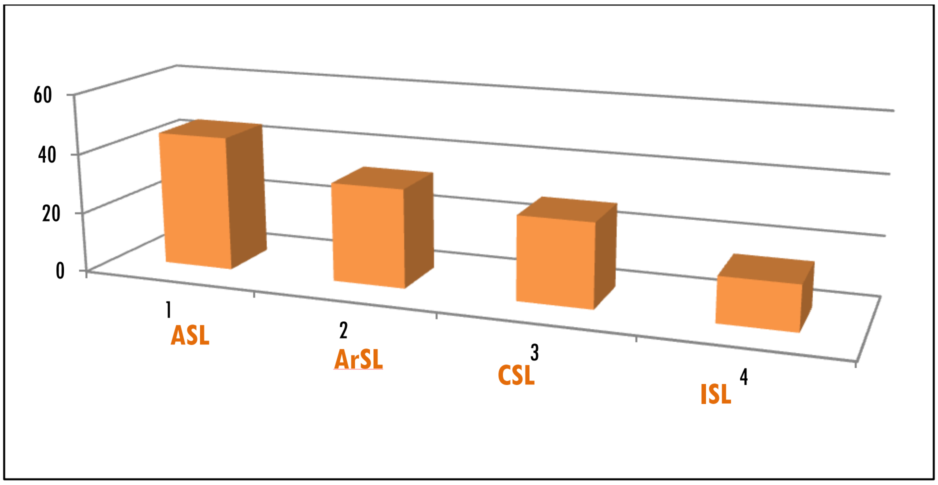



**Figure 2.4:** Volume of research done in various Sign Languages

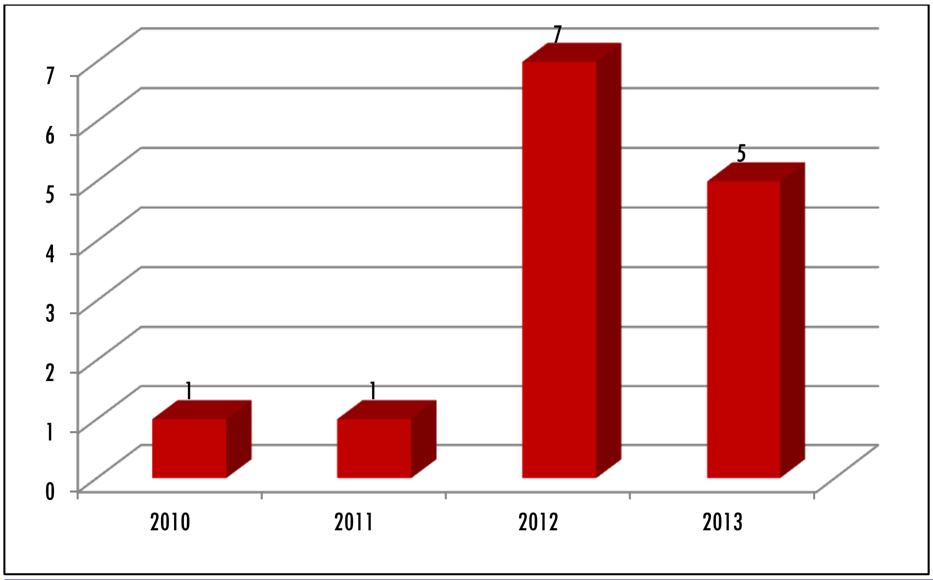



**Figure 2.5:** Volume of research work done in ISL

Sign language recognition involves simultaneous monitoring of different body articulators and their synchronisation and integration by following a multimodal approach [1]. Three main channels that require focus in sign language recognition are static hand gestures where hand shape/pose represents a particular meaning, dynamic gestures consisting of hand shape and motion trajectory and facial expressions. Several complexities are associated with these channels and the performance of the recognition system depends significantly on the way in which there complexities are addressed. Therefore this work concentrates on finding decisive feature extraction methods which can help in the building of high performance ISL recognition system. The constraints that make gesture recognition complex are:

1. Static gestures with resemblances
2. Static overlaid gestures
3. Similar dynamic gestures giving different meaning depending on the hand motion trajectory.
4. Facial expression changes occurring in sign language sentences.

Chapter 3, 4 and 5 discuss in detail about the selection of appropriate feature descriptors for the three main channels of sign language communication taking into account the identified complexities.

## 2.5 Summary of the Chapter:

The literature survey enabled to identify the quantum of research work carried out in various sign languages and their success. Many research works are carried out in ASL covering different channels of sign language recognition. Research works in Arabic and Chinese sign languages are focused mainly on manual gestures. The survey also indicated the paucity of works in facial expression changes and integrating different channels of sign language recognition. Each sign language is different and therefore, independent research is required due to their inherent complexities. Therefore, this study conveys the need to conduct more research in ISL and the requirement to test feature descriptors for their potential to tackle the hidden complexities in gesture recognition.

……….cs80……….

*Chapter* **3**

# STATIC HAND GESTURE RECOGNITION USING SEMI GLOBAL DESCRIPTORS

Contents



*Hand gestures with resemblances and overlaid hand gesture patterns are the two main complexities encountered in static hand gestures. This chapter evaluates the different feature descriptors for their potential to identify complex static hand gestures. The analysis provided the best feature descriptor for hand shape identification.*

## 3.1 Introduction

The overall analysis of selected reviews clearly indicated the advancement of sign language recognition research globally and on Indian context. Apart from few promising works, most of the research works in ISL recognition system were on static gestures. In addition to static gestures, ISL also consists of dynamic gesture and non-manual gestures, where less research has been done in Indian context. Hence our research work concentrated on different channels in ISL and tried to focus and address the associated complexities. Analysis results identified that some feature descriptors or their combination may yield higher accuracy and recognition rate while handling complex gestures in ISL.

Static hand gestures represent different hand postures that will not vary over time. Hand postures differ each other due to the difference in the projection of fingers over palm area, bending of fingers, placing of hand or fingers on different body parts with a particular posture. However, there can be similarity in gestures.

Figure 3.1 shows visual similarity in gesture representation for letters D, L and I due to the projection of a single finger over palm area at different positions [67]. The relative positioning of this finger to palm is a crucial aspect and a gesture recognition system needs to account this complexity. In finger 3.2 overlaid positions, number of fingers sign for different letters. For example, gestures M and N depend on the number of fingers placed over the palm area. Gestures for 'MOTHER', 'THINK' and 'CONFUSE' differ from each other based on the position of finger on face (Figure 3.3). In addition to similarity, the overlaid hand gestures encounter more complexity for recognition as base and overlapped areas have the same texture.


**Figure 3.1:** Hand gestures with resemblance

**Figure 3.2:** Hand overlaid on palm for M and N gestures

**Figure 3.3:** Hand on face area for MOTHER, THINK and CONFUSE

As explained, gestures with resemblances and overlaid hand gestures offer increased complexities in ISL static gesture recognition. However, by selecting appropriate feature descriptors, these complexities can be minimized. Selection of feature descriptors for images is based upon the global, semi global and local level [68] visualization of them. Global descriptor views image as a whole for feature extraction and local descriptor generates features based only on the prominent points in that image, and a semi global descriptor evaluates an image in a sub-regional level.

Histograms of Orientated Gradients (HOG) feature descriptor method is widely used in human detection and object recognition [69]. HOG's capability to evaluate an image at its sub regional levels, make it suitable for the identification of complex hand gestures.

Local Binary Pattern (LBP) [71] feature descriptor is well known for its highly discriminative capacity for texture recognition. The texture discriminative potential of LBP feature descriptor makes it suitable for complex gesture recognition process.

In this study, a comparison of HOG and LBP with other well-known feature descriptors like SIFT, SURF, HU moment invariant, PCA used for image recognition was also done.

## 3.2 Recognition Framework

The hand gesture recognition system generally consists of training and testing phases. The system architecture of a static hand gesture recognition scheme is illustrated in Figure 3.4.

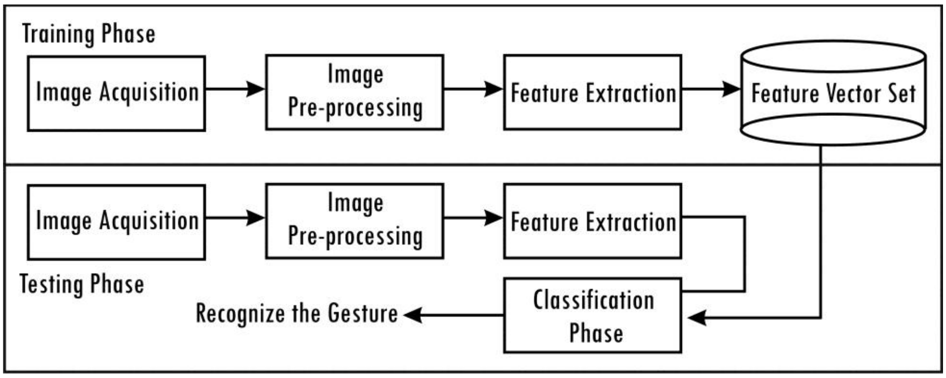



**Figure 3.4:** System Architecture

### 3.2.1 Image Pre-processing

Bare hand gestures are captured for hand detection and segmentation process. From this image, hand region is extracted from the background based on the skin color. Skin pixel region is identified from other color regions by using a threshold in RGB color space, where the threshold values of R, G and B are selected experimentally using a set of rules [72] as given in Table 3.1.

**Table 3.1:** Set of rules for the classify a pixel as skin color

| |
|---|
| A pixel is classified as skin color only if it's R, G and B values are: |
| R > 50, G > 40, B > 20 |
| Max(R, G, B) – Min(R, G, B) > 15 |
| \|R – G\| > 15 |
| R < G |
| R > B |

By this approach, skin color pixels are extracted from the non-skin color pixels. The resultant image is subjected to connected component analysis to obtain the segmented hand region. Segmented RGB of hand image is then converted into normalized gray scale image. Figure 3.5 shows the various stages.

**Figure 3.5:** Image segmentation based on skin color from background

## 3.3 Feature Descriptors

Achievement of high recognition rate indicates the superiority of a feature extraction method employed in a system. Selection of feature descriptors is the essence of a recognition system. The different schemes used for extracting discriminative features are presented below.

### 3.3.1 Histograms of Orientation Gradients (HOG)

HOG is a semi global descriptor that can evaluate an image in sub regional level [69].This method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid [67]. It is processed as follows.

1. Obtain the normalized gray scale image
2. Divide images into cells
3. Gradient magnitude (G)and orientation (θ) are computed for all pixels in cells using the Eq 3.1 and Eq 3.2

$$|G| = \sqrt{I_x^2 + I_y^2} \qquad (3.1)$$

$$\theta = \tan^{-1}\frac{I_x}{I_y} \tag{3.2}$$

$I_x$ and $I_y$ are horizontal and vertical gradients of the image *I*, obtained by using a convolution operation as given by Eq3.3.

$$I_x = I*D_x, \text{ and } I_y = I*D_y \tag{3.3}$$

Where $D_x$ and $D_y$ are 1D-filter kernels,

$$Dx = [-1\ 0\ 1] \text{ and } Dy = [1\ 0\ -1]^T$$

4. Each pixel within the cell casts vote to an orientation based histogram bin corresponding to the values found in the gradient magnitude computation. The histogram bins are evenly spread over $0^0$ to $180^0$, as shown in Fig 3.6.

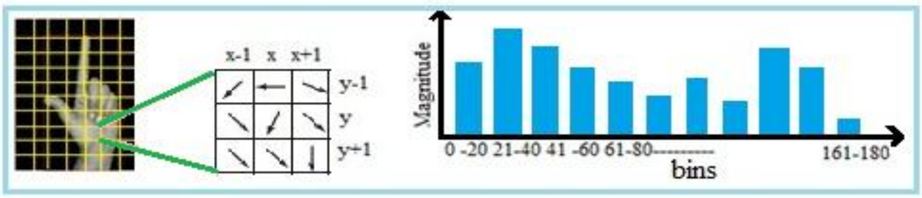



**Figure 3.6:** Image divided into cells and corresponding histogram representation.

5. In order to account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into overlapped blocks as shown in Figure 3.7. All normalized block histograms are concatenated to form the entire HOG feature vector using probability density function.

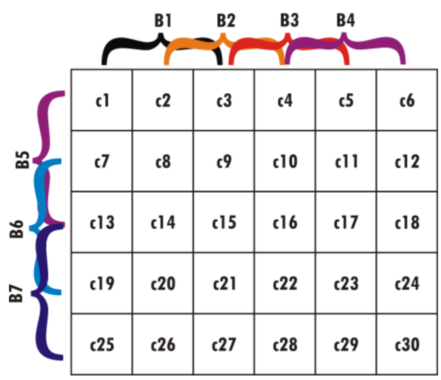



**Figure 3.7:** HOG divides an image into cells and blocks for feature extraction.

### 3.3.2 Local Binary Pattern (LBP)

LBP is used to extract feature vector from image cells based on the pixel value difference. It is calculated as follows.

1. Obtain the gray scale image
2. Divide the image into cells
3. LBP value for the central pixel of the cell can be obtained by:

$$LBP_{P,R} = \sum_{P=0}^{P-1} s\,(g_p - g_c)2^p \tag{3.4}$$

Where $s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$

$g_c$ is the gray value of the central pixel

$g_p$ is the value of its neighbours

P is the total number of neighbours involves

R is the radius of the neighbourhood

4. Compute the histogram over the block based on the occurrence of LBP values
5. A final histogram is obtained by concatenating histograms of all cells and then normalized.

This is considered as the feature vector corresponding to an image.

## 3.4 Dataset

The hand gestures are collected under non uniform background with different illumination and from different age groups. The training dataset contains images collected from five signers who did ten replications of a sign at different times (Table 3.1). For testing, different dataset was collected from two signers who did ten samples for each gesture at different time. Static hand gesture recognition was done on four datasets. They include dataset representing single hand alphabets, overlaid hand on hand gestures, overlaid hand on face gestures and a set of double hand ISL alphabets as shown fig 3.8, 3.9, 3.10, 3.11.

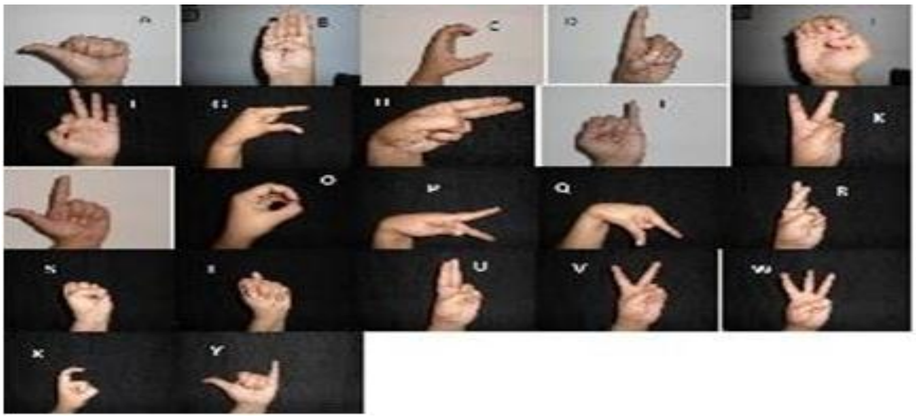


**Figure 3.8:** Single hand alphabets

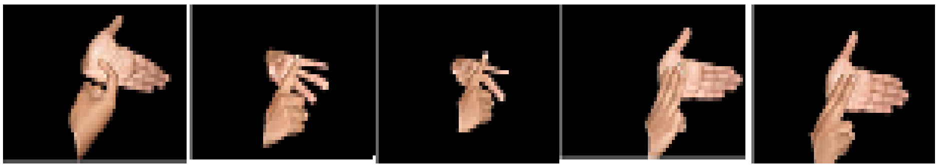


**Figure 3.9:** overlaid hand on hand.

**Figure 3.10:** Overlaid hand on face

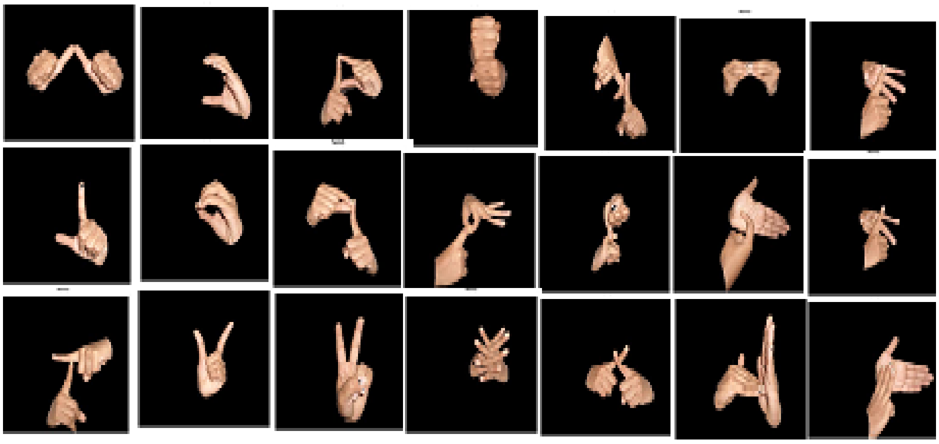


**Figure 3.11:** Double hand ISL alphabets.

**Table 3.2**: Dataset details

| Dataset | No. of Classes | Samples in Training set | Samples in Testing set |
|---|---|---|---|
| Single hand ISL alphabets | 22 | 1,100 | 440 |
| Overlaid hand on hand | 5 | 250 | 100 |
| Overlaid hand on face | 5 | 250 | 100 |
| Double hand ISL alphabets | 21 | 1,050 | 420 |

## 3.5 Implementation

Static hand gesture recognition system consists of training phase and testing phase (Fig 3.4). Four training dataset and four testing dataset were used by this system (Table 3.2).

**Training Phase**

In this phase RGB color images in the training dataset are resized to image of resolution 320 x 240. This image is given to segmentation module in order to extract the hand region based on skin color. It is then converted to gray scale image. HOG features were extracted from single hand alphabet set, and from two overlaid gesture sets. LBP features were extracted from two overlaid gesture sets. Combined HOG-LBP features were generated from two overlaid hand gesture sets and from are ISL double hand alphabet set. Total six feature vector sets were generated in the training phase.

**Testing Phase**

In this phase, feature vector corresponding to the test gesture is extracted. Then the classifier recognizes the test gesture based on the trained feature vector set.

## 3.6 Experimental results and Performance evaluation

Statistical measures like precision, recall/sensitivity, specificity, F-measure and accuracy are commonly used for evaluating the accuracy of any classifier model. The corresponding values obtained for single hand alphabet dataset through the experiment are tabulated in Table 3.8 and confusion matrix of each method is shown in Table 3.3, 3.4, 3.5, 3.6 & 3.7. A comparison of these values across the different feature extraction methods is shown in Fig 3.12. Selection of feature descriptors in those feature extraction methods considered global, semi global and local level visualization of the images. Scale Invariant Feature Transform SIFT [73] and Speeded Up Robust Feature method SURF [74] consider from local visualization while Hu Moment Invariant [75] and Principal Component Analysis (PCA) [76] consider global visualized. An experiment was also carried out combining Hu Moment and SURF descriptors. From the table it can be seen that for single hand alphabets with resemblances HOG descriptors and K-NN classifier with k value as 5 gave the highest rate of accuracy (96%). The superiority of HOG based method can be attributed to the fact that HOG extracts features at sub region level of the image, preventing lose of any prominent features of the gesture.

In overlaid hand gestures, complexity due to textural similarity is encountered in addition to resemblances of gestures. The experimental results underline the superiority of semi global descriptors in handling overlaid gestures with resemblances. LBP, a highly discriminative semi global texture descriptor was experimented for identifying static overlaid gestures. Evaluation was done using two datasets - hand overlaid on hand gestures and hand overlaid on face gestures. The results from this analysis are presented in Table 3.9, 3.10, 3.11, 3.12 , 3.13 and 3.14. Fig 3.13 and 3.14 gives a comparison of various methods used for handling overlaid gestures. The analysis indicates that the HOG-LBP combination has better performance over the other individual methods. The accuracy of HOG on handling hand on hand gestures is slightly higher than LBP due to the specific potential of it in differentiating gestures with resemblances. However, best results were obtained when these methods were combined. The comparative analysis also showed better results for SVM classifier over K-NN. The combination of HOG-LBP features and SVM classifier gave an accuracy rate of 93.8% for ISL double hand alphabet database as shown in Fig 3.15.

Graphical representation in Fig 3.15 shows a concomitant increase for specificity and sensitivity values. The same trend was also observed for precision and recall, this clearly indicates the better level of performance of the recognition system. In addition, the average values of all the measures were above 0.83. Therefore, the analysis clearly indicates the suitability of combined HOG-LBP (semi global descriptor which evaluates images at sub regional level) method for static hand gesture recognition

**Table 3.3:** Confusion Matrix for Single Hand Alphabet Recognition System Using SURF Feature Extraction Method.

| | A | B | C | D | E | F | G | H | I | K | L | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.60 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| B | 0.10 | 0.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C | 0.10 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| D | 0.00 | 0.00 | 0.20 | 0.70 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | 0.60 | 0.00 | 0.00 | 0.00 | 0.20 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.20 | 0.00 | 0.00 |
| G | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.50 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| H | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.00 | 0.00 | 0.00 | 0.20 | 0.15 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| K | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 |
| L | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.15 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| O | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| P | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| S | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.90 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| T | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 |
| V | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 |
| W | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 |
| X | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 |
| Y | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 |

**Table 3.4:** Confusion Matrix for Single Hand Alphabet Recognition System Using SIFT Feature Extraction Method.

| | A | B | C | D | E | F | G | H | I | K | L | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.60 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 |
| B | 0.00 | 0.95 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 |
| D | 0.00 | 0.00 | 0.00 | 0.50 | 0.30 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | 0.60 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.20 | 0.00 | 0.00 | 0.00 |
| G | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.45 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| H | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| K | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 |
| L | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| O | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| P | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| S | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 |
| T | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.30 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 |
| V | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 |
| W | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 |
| X | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 |
| Y | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 |

**Table 3.5:** Confusion Matrix for Single Hand Alphabet Recognition System Using PCA Feature Extraction Method.

| | A | B | C | D | E | F | G | H | I | K | L | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.65 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| B | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 |
| D | 0.00 | 0.00 | 0.10 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 |
| G | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| H | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| K | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.40 | 0.00 | 0.00 | 0.00 | 0.00 |
| L | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| O | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 |
| P | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.65 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 |
| S | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 |
| T | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 |
| V | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 | 0.00 |
| W | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.85 | 0.00 | 0.00 |
| X | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 |
| Y | 0.05 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 |

**Table 3.6:** Confusion Matrix for Single Hand Alphabet Recognition System Using Hu-moment Feature Extraction Method.

| | A | B | C | D | E | F | G | H | I | K | L | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.65 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 |
| B | 0.00 | 0.95 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| D | 0.00 | 0.00 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.20 | 0.00 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 |
| G | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| H | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.55 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| K | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.00 | 0.00 | 0.00 |
| L | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 |
| O | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| P | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.45 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.70 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| S | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| T | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 |
| V | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 |
| W | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 |
| X | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 |
| Y | 0.10 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.70 |

**Table 3.7:** Confusion Matrix for Single Hand Alphabet Recognition System Using HOG Feature Extraction Method.

| | A | B | C | D | E | F | G | H | I | K | L | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.96 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| B | 0.00 | 0.96 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| D | 0.00 | 0.00 | 0.00 | 0.92 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| E | 0.00 | 0.00 | 0.00 | 0.00 | 0.92 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 |
| G | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| H | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.92 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 |
| K | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.98 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 |
| L | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.97 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| O | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| P | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.97 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.97 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.92 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 |
| S | 0.03 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| T | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.82 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.96 | 0.00 | 0.00 | 0.00 | 0.00 |
| V | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.00 | 0.00 |
| W | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 | 0.00 | 0.00 |
| X | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| Y | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 |

**Table 3.8:** Performance measures of K-NN classifier with different feature extraction methods for recognizing hand gestures with resemblances

| Feature Extraction Methods | Specificity (True Negative Rate) | Sensitivity/ Recall (True Positive Rate) | Precision (Positive Predictive Value) | F-measure | Accuracy |
|---|---|---|---|---|---|
| PCA | 0.93 | 0.48 | 0.5 | 0.45 | 0.88 |
| SIFT | 0.91 | 0.59 | 0.5 | 0.51 | 0.88 |
| SURF | 0.91 | 0.44 | 0.46 | 0.44 | 0.84 |
| Hu moment invariant | 0.92 | 0.57 | 0.61 | 0.58 | 0.89 |
| SURF-moment | 0.92 | 0.57 | 0.64 | 0.58 | 0.87 |
| HOG | **0.96** | **0.86** | **0.80** | **0.83** | **0.96** |

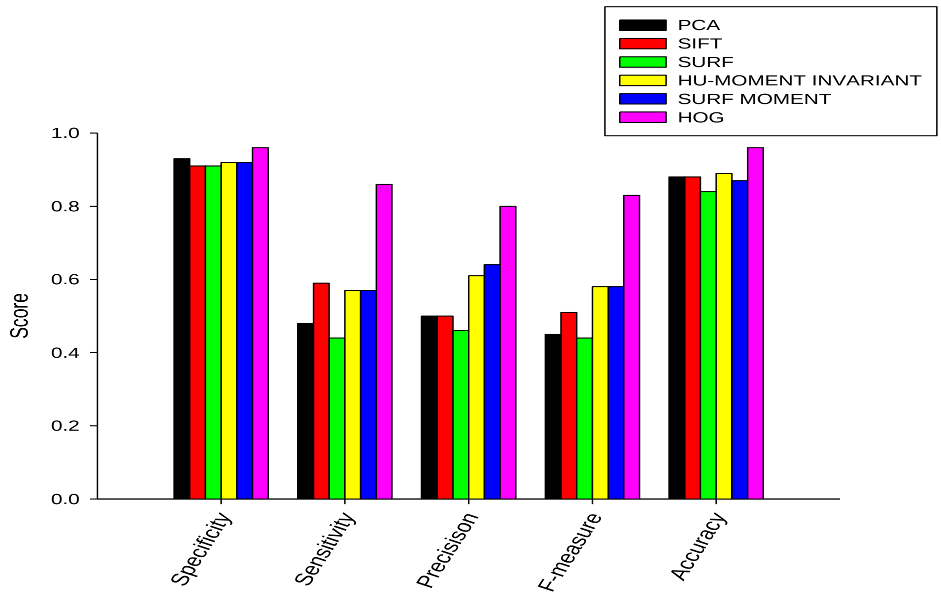
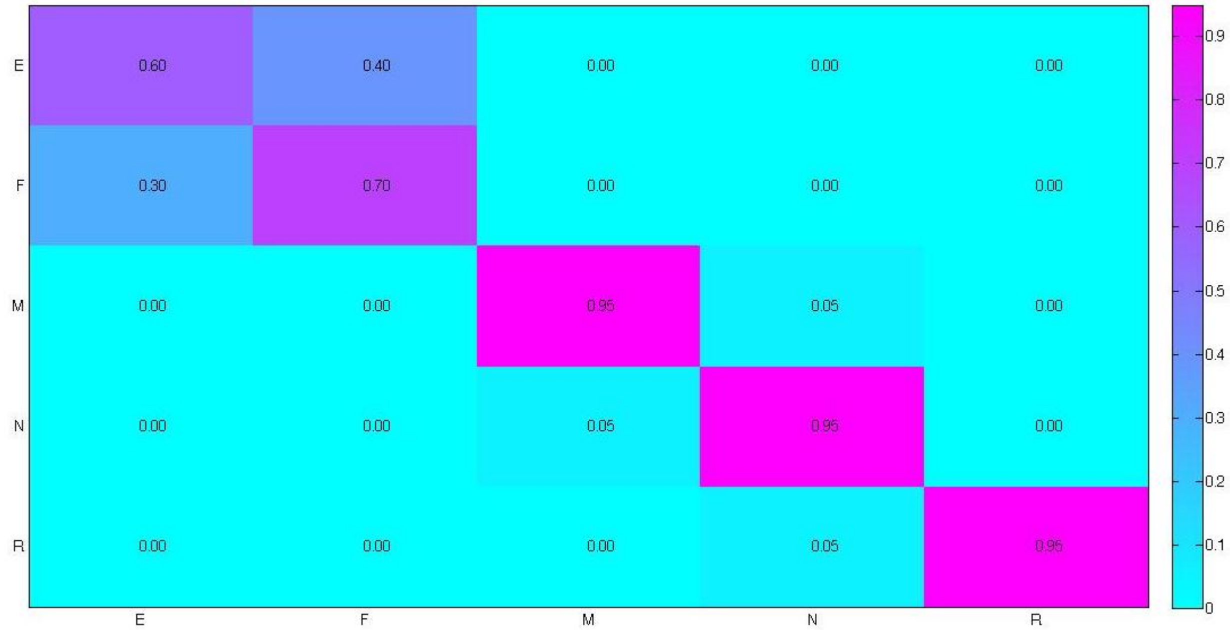



**Figure 3.12:** A comparison of methods for handling hand gestures with resemblances

**Table 3.9:** Confusion Matrix for Overlaid Hand on Hand Recognition System Using HOG Feature Extraction Method.

| | E | F | M | N | R |
|---|---|---|---|---|---|
| E | 0.95 | 0.05 | 0.00 | 0.00 | 0.00 |
| F | 0.04 | 0.96 | 0.00 | 0.00 | 0.00 |
| M | 0.00 | 0.00 | 0.65 | 0.35 | 0.00 |
| N | 0.00 | 0.00 | 0.45 | 0.55 | 0.00 |
| R | 0.00 | 0.00 | 0.15 | 0.10 | 0.75 |

**Table 3.10:** Confusion Matrix for Overlaid Hand on Hand Recognition System Using LBP Feature Extraction Method.



| | E | F | M | N | R |
|---|---|---|---|---|---|
| E | 0.60 | 0.40 | 0.00 | 0.00 | 0.00 |
| F | 0.30 | 0.70 | 0.00 | 0.00 | 0.00 |
| M | 0.00 | 0.00 | 0.95 | 0.05 | 0.00 |
| N | 0.00 | 0.00 | 0.05 | 0.95 | 0.00 |
| R | 0.00 | 0.00 | 0.00 | 0.05 | 0.95 |

**Table 3.11:** Confusion Matrix for Overlaid Hand on Hand Recognition System Using Combined HOG and LBP Feature Extraction Method with KNN classifier.

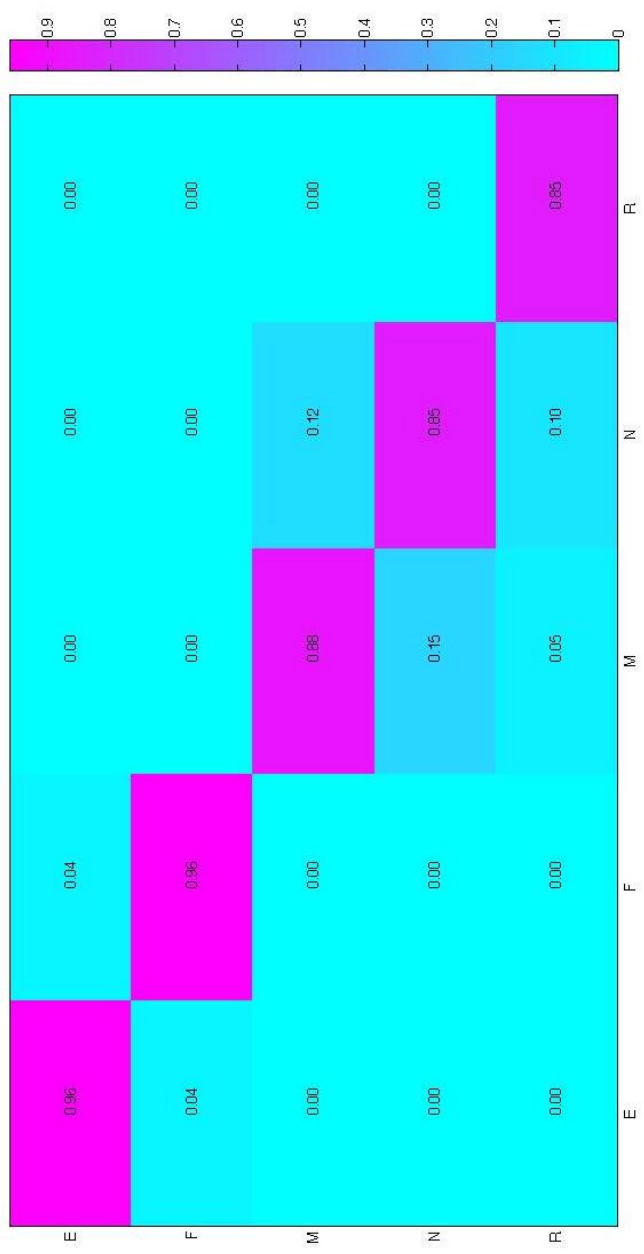


**Table 3.12:** Confusion Matrix for Overlaid Hand on Hand Recognition System Using Combined HOG and LBP Feature Extraction Method with SVM classifier.

| | E | F | M | N | R |
|---|---|---|---|---|---|
| E | 0.96 | 0.04 | 0.00 | 0.00 | 0.00 |
| F | 0.04 | 0.96 | 0.00 | 0.00 | 0.00 |
| M | 0.00 | 0.00 | 0.90 | 0.10 | 0.00 |
| N | 0.00 | 0.00 | 0.10 | 0.90 | 0.00 |
| R | 0.00 | 0.00 | 0.05 | 0.05 | 0.90 |

**Table 3.13:** Performance measures of classifiers for Hand-on-Hand gestures.

| | Precision | Recall /sensitivity | Specificity | F_measure | Accuracy |
|---|---|---|---|---|---|
| LBP+SVM | 0.55 | 0.71 | 0.79 | 0.62 | .83 |
| HOG+SVM | 0.79 | 0.65 | 0.89 | 0.71 | .86 |
| HOG-LBP+KNN | 0.81 | 0.79 | 0.89 | 0.8 | .89 |
| HOG-LBP+SVM | 0.89 | 0.82 | 0.91 | 0.85 | .93 |

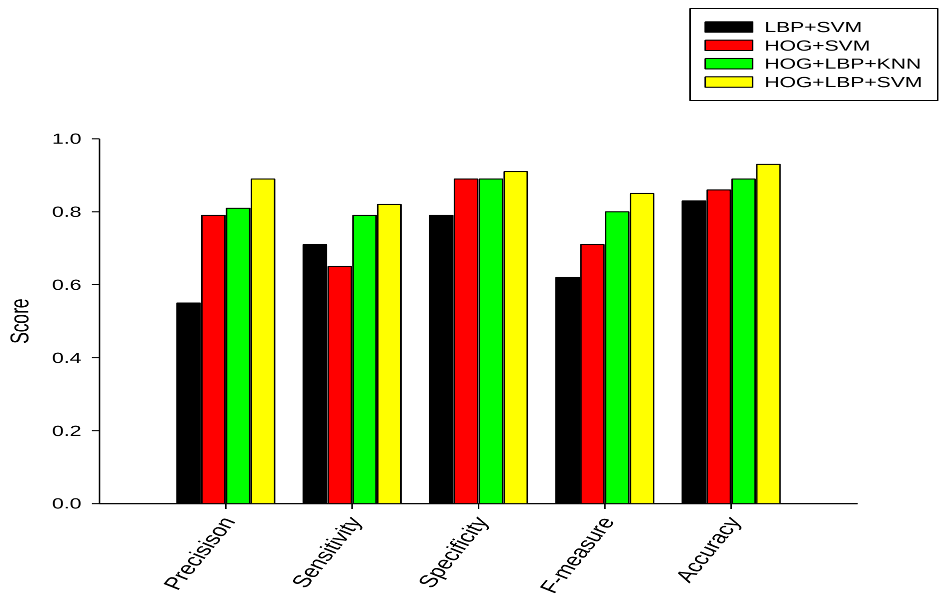



**Figure 3.13:** Graphical representation of performance measures given in Table 3.13

**Table 3.14:** Performance measures of classifiers for Hand-on-Face gestures.

| Methods | Precision | Recall/ sensitivity | Specificity | F_measure | Accuracy |
|---|---|---|---|---|---|
| LBP+SVM | 0.6 | 0.7 | 0.8 | 0.64 | 0.86 |
| HOG+SVM | 0.5 | 0.6 | 0.78 | 0.55 | 0.81 |
| HOG-LBP+KNN | 0.7 | 0.8 | 0.82 | 0.74 | 0.89 |
| HOG-LBP+SVM | 0.83 | 0.9 | 0.96 | 0.81 | 0.91 |

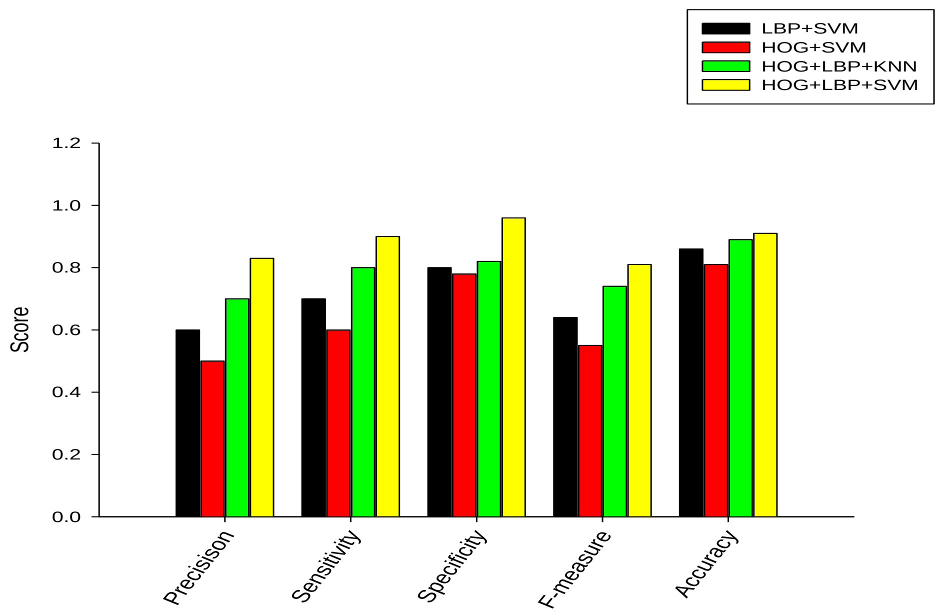
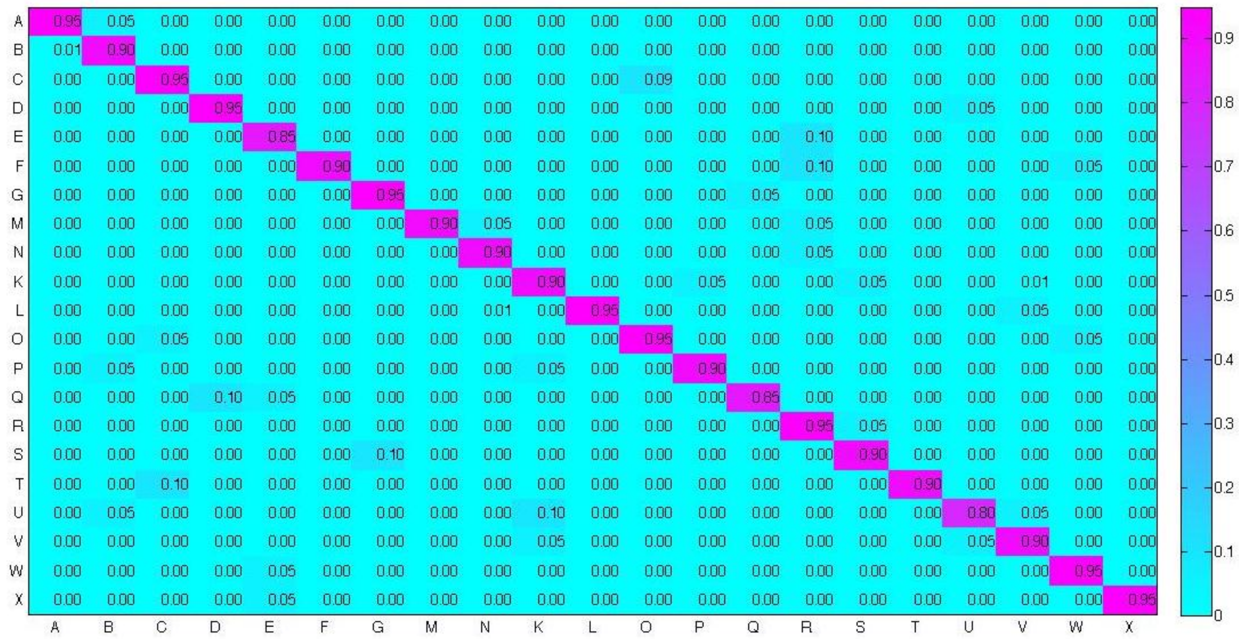



**Figure 3.14:** Graphical representation of performance measures given in Table 3.14

**Table 3.15:** Confusion Matrix for Double Hand ISL Alphabet Recognition System Using Combined HOG and LBP Feature Extraction Method with SVM classifier



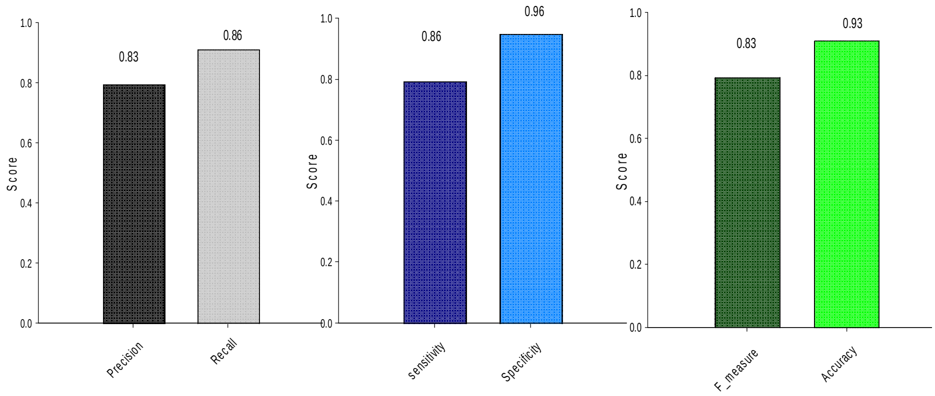



**Figure 3.15:** Average values of performance measures of static hand gesture recognition system on double hand ISL alphabet database.

## 3.7 Summary of the Chapter

Gestures with resemblances and overlaid hand gestures offers increased complexities in ISL static gestures. Global, semi global and local feature descriptors were considered for the static hand gesture recognition of which, combined HOG-LBP (semi global feature descriptors), gave the best performance.

Chapter 4

# HAND MOTION TRAJECTORY ANALYSIS FOR DYNAMIC HAND GESTURE RECOGNITION USING ORIENTATION FEATURES

Contents



*Dynamic hand gestures are intrinsic component in sign language communication. Extracting spatial and temporal features of the hand gesture trajectory, plays an important role in dynamic gesture recognition. Finding a discrete feature descriptor for the motion trajectory based on the orientation feature is the main concern of this chapter. Kalman filter algorithm and Hidden Markov Models (HMM) models are used in the system.*

## 4.1 Introduction

A hand shape along with its movement is considered as dynamic hand gesture. Signing of alphabet J is dynamic and various stages of its representation is shown in Fig 4.1. This example clearly illustrates the spatial and temporal changes occurring to dynamic gesture frames and the need to correctly explore those changes to have recognition accuracy.

The spatial and temporal characteristic changes make dynamic gesture recognition a difficult task. From the above example, it is clear that hand shape detection and motion sequence recognition are the two components which contribute to the dynamic gesture recognition. Hence the important steps in dynamic hand gesture recognition are (1) hand shape recognition, (2) hand motion tracking, (3) classification for spatial and temporal characterization.

Extracting discrete feature vector from trajectory motion plays an important role in this recognition system. Location, orientation and velocity are the basic three features of a hand motion trajectory. Among these three features, orientation feature is very important for having high recognition accuracy [80,81,82]. Therefore, this study focuses on orientation feature as a main feature for dynamic gesture recognition.

Path of the motion needs to be tracked for the extraction of discrete features. Kalman filter algorithm [77] is widely used for generating the trajectory paths in many applications like robotic motion control and moving object detection. This method has been used successfully for hand motion tracking [78, 79].

Hidden Markov Model (HMM) is a statistical model widely used in hand writing, speech and character recognition because of its capability of modelling spatial- temporal time series [84].In this study, training and testing of gesture path is done by HMM as it has the capability to preserve the spatial –temporal identity of the motion path.

**Figure 4.1:** Dynamic gesture representing alphabet 'J'

## 4.2 Recognition Framework

A general framework of a dynamic hand gesture recognition system is shown in Fig 4.2.System consists of two main stages and they are hand shape recognition and hand motion trajectory recognition. A classifier correctly identifies the dynamic hand gesture. This chapter concentrates on the extraction of the discrete feature vector of the gesture trajectory path.

### 4.2.1Pre-processing

Input to the dynamic hand gesture recognition system is a video stream. Extraction of frames from the video input is the main operation in the pre-processing phase. The first frame of this sequence is passed to the hand shape recognizer for the hand shape identification of the gesture while total frame sequences are passed to the motion trajectory recognizer.

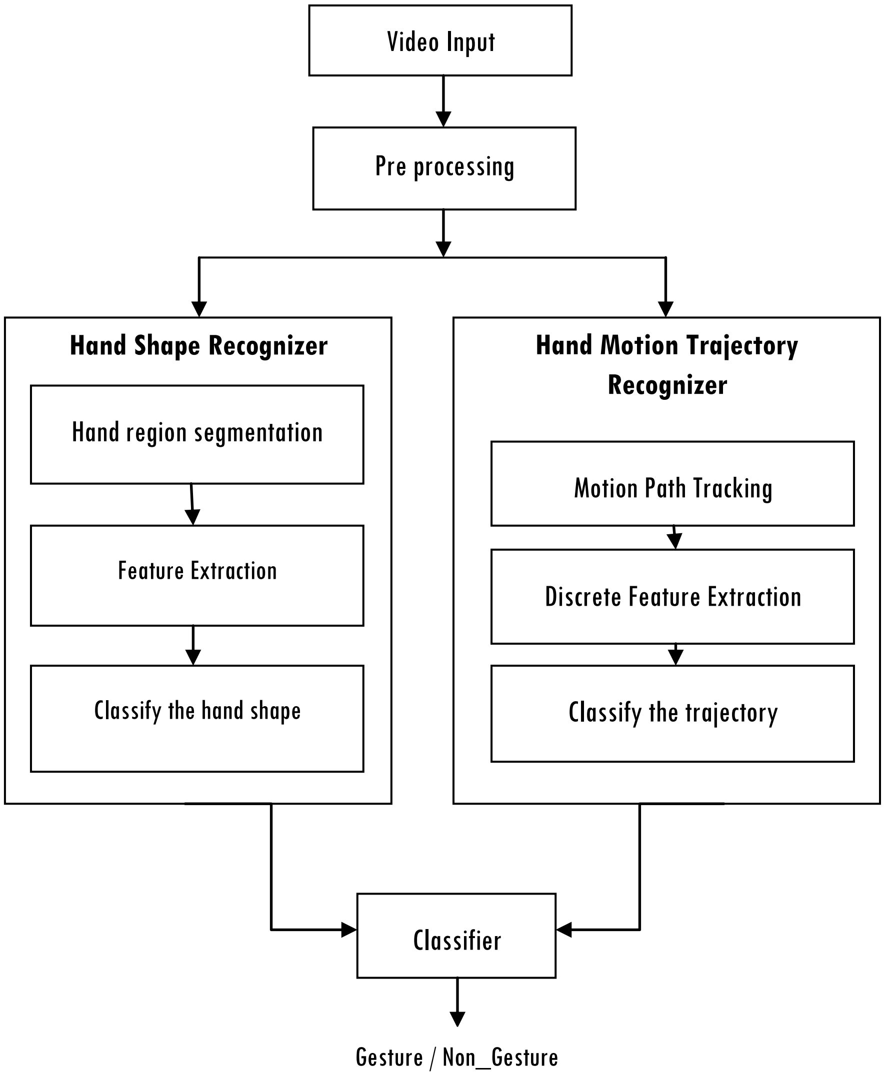



**Figure 4.2:** Framework of a dynamic hand gesture recognition system.

### 4.2.2 Hand Shape and Hand Trajectory Motion Recognizer

#### 4.2.2.1 Hand Shape Recognizer

Shape of the gesture would be recognized from the first frame of the image sequence. Hand area is identified based on skin color, and the area is segmented for feature vector extraction. Hand shape recognizer follows the same method as explained in Chapter 3.

#### 4.2.2.2 Hand Trajectory Motion Recognizer

The hand is tracked from each frame to generate the motion trajectory path. The centroid of hand is identified from each frame and it is used for tracking the trajectory path of the hand motion. Discrete feature vector is extracted for this path and is used for identifying the hand gesture.

## 4.3 Discrete Feature Vector for Gesture Path

This phase plays a key role in dynamic gesture recognition. Change in the direction of hand movement during gesturing is the orientation feature of the hand motion. Hand motion path consists of a series of centroid points of the moving hand represented by Cartesian coordinates. Orientation feature of the motion trajectory can be represented as the angle changes between the centroid points. These angle changes are considered as the feature vector of the trajectory motion. Selection of the centroid points to generate these angles has a significant role in motion trajectory analysis.

In this chapter two feature vectors are extracted (1) orientation angle change ($\theta_1$),(2) directional angle change ($\theta_2$). The orientation angle $\theta_1$is calculated by taking two consecutive centroid points, with respect to horizontal plane while the directional angle change $\theta_2$ is computed by considering three consecutive centroid points in the trajectory path as shown in Fig 4.3.

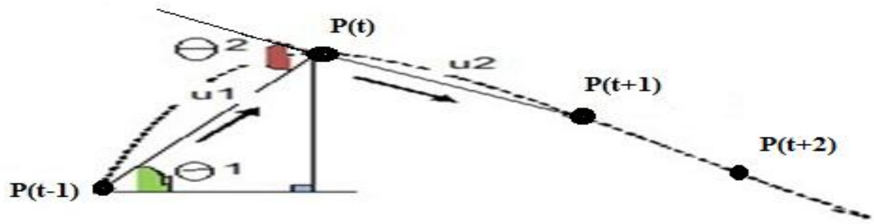



**Figure 4.3** Trajectory paths with consecutive points P (t-1), P(t),P(t+1) and the two angle positions.

For the centroid point P (t) = $(x_t, y_t)$; t =1,2,……T, where T is the length of the motion path, the orientation angle change ($\theta_1$) and directional angle change ($\theta_2$) are computed using the formula(4.1) and (4.2).

$$\theta_1 = \arctan((y_{t-1} - y_{t)})/((x_{t-1} - x_{t)}) \qquad (4.1)$$

$$\theta_2 = \text{cosinv}\frac{u1 \, .u2}{|u1|| \, u2|} \qquad (4.2)$$

where P (t) = $(x_t, y_t)$

P (t-1) = $(x_{t-1}, y_{t-1})$

P (t+1) = $(x_{t+1}, y_{t-1})$

$u_1$ =P (t+1) – P (t)

$u_2$ = P (t+2) – P (t+1)

Two discrete feature vector f ($\theta_1$) and f ($\theta_2$) is extracted from the gesture path corresponding to the orientation angle θ1 and directional angle θ2 respectively.

## 4.4 Dataset

Two video datasets were collected for analysis. The first set represented numbers 0-9 by moving the hand in space, as shown in

Fig4.4. The second set was dynamic hand gestures in ISL as shown in Fig 4.5. Each video was of length 5 secs. Five signers were selected to create the training dataset. Each signer did ten replications on a single sign at different times. Size of the training dataset was 500 while the test data of size 200, (Table 4.1).

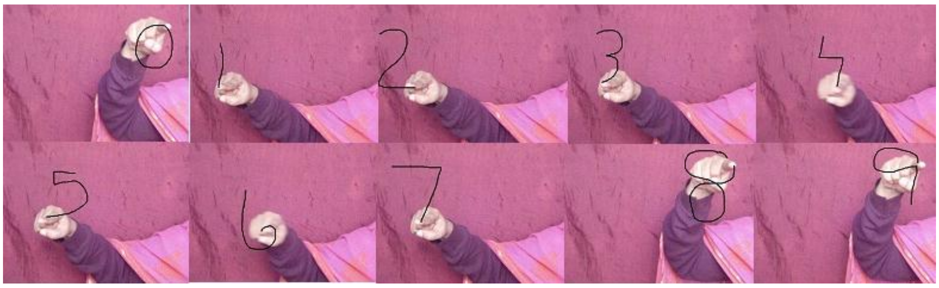


**Figure 4.4:** Represent numbers 0-9 by moving the hand

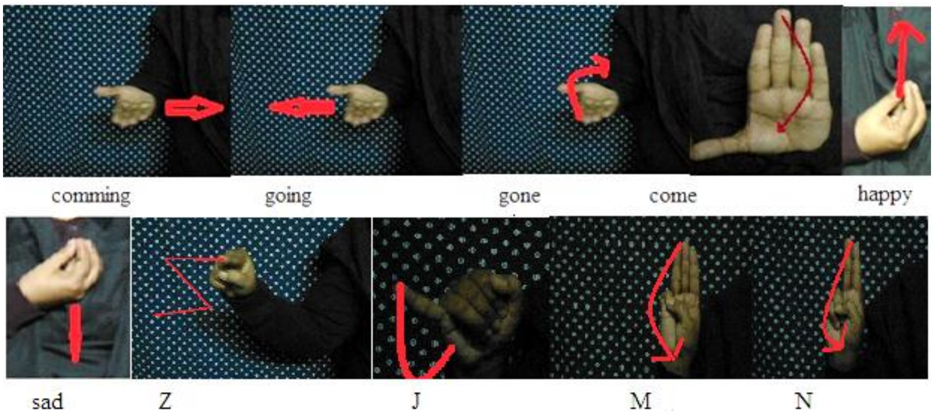



**Figure 4.5:** Dynamic Hand Gestures in ISL.

**Table 4.1:** Dataset details

| Dataset | No. of Classes | Samples in Training set | Samples in Testing set |
|---|---|---|---|
| Representing numbers 0-9 by moving the hand in space. | 10 | 500 | 200 |
| Dynamic hand gestures in ISL | 10 | 500 | 200 |

## 4.5 Implementation

The system consists of mainly two phase, training phase and testing phase.

**Training Phase**

The frames are extracted from the video input at the pre-processing stage. In the training phase, hand shape recognizer generates feature vector set for hand gestures based on the first frame extracted from the video inputs.

The video frame sequences are then fed to hand motion trajectory recognizer. The motion path (gesture path) is tracked from these frames, as shown in fig 4.6. Identifying the moving hand is done by taking the image difference between the consecutive frames and finding the centroid point of this area using Kalman filter algorithm. Such centroid points in fifteen equidistance frames in a video are consider for path tracking. Two discrete feature vectors f ($\theta_1$) and f ($\theta_2$) are extracted based on the orientation of the trajectory path. LR based HMM is used for modelling the trajectory and these discrete feature vectors are used as the observation sequences for it. One HMM is designed for each gesture. Number of states in each HMM will vary according to the directional changes in the trajectory of the gesture, as shown in Figure 4.7. Each digit has a separate HMM. HMMs for all the digits 0 to 9 are given in Figure 4.8.

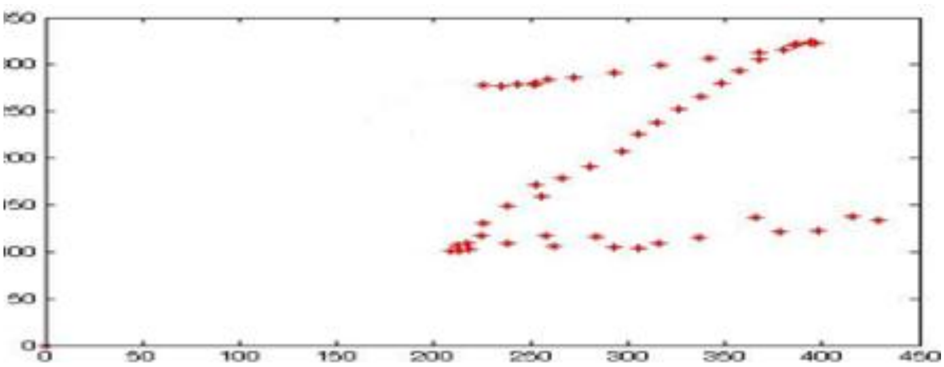



**Figure 4.6:** Motion trajectory for alphabet Z.

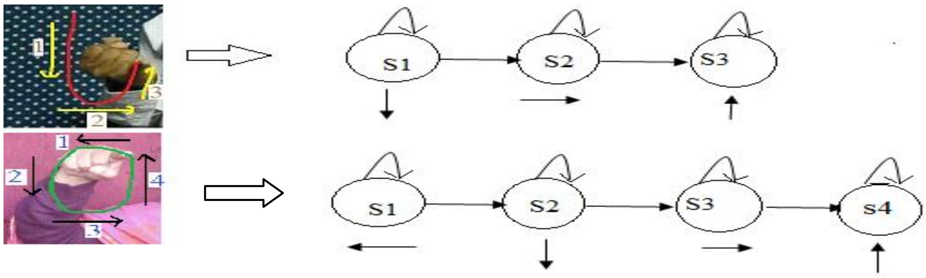



**Figure 4.7:** Motion path and corresponding states in HMM

The HMM trajectory models for each gesture is trained using Baum-Welch algorithm [87]. HMM parameters are computed for each type of gesture during the training process.

**Testing Phase**

In this phase, the features of the gestures to be recognized are extracted. HOG-LBP feature vector corresponding to hand shape is extracted and SVM classifier is used for identifying the hand shape. Discrete feature vectors generated from the test gesture motion path are feed to the HMM model set. HMM model set will identify the motion trajectory class for the test gesture as shown in Fig 4.9. Gesture class identified by hand shape recognizer and motion trajectory recognizer was compared. If both were identified as to the same class, then the tested dynamic hand gesture is classified as to that class, otherwise treat it as non gesture type.

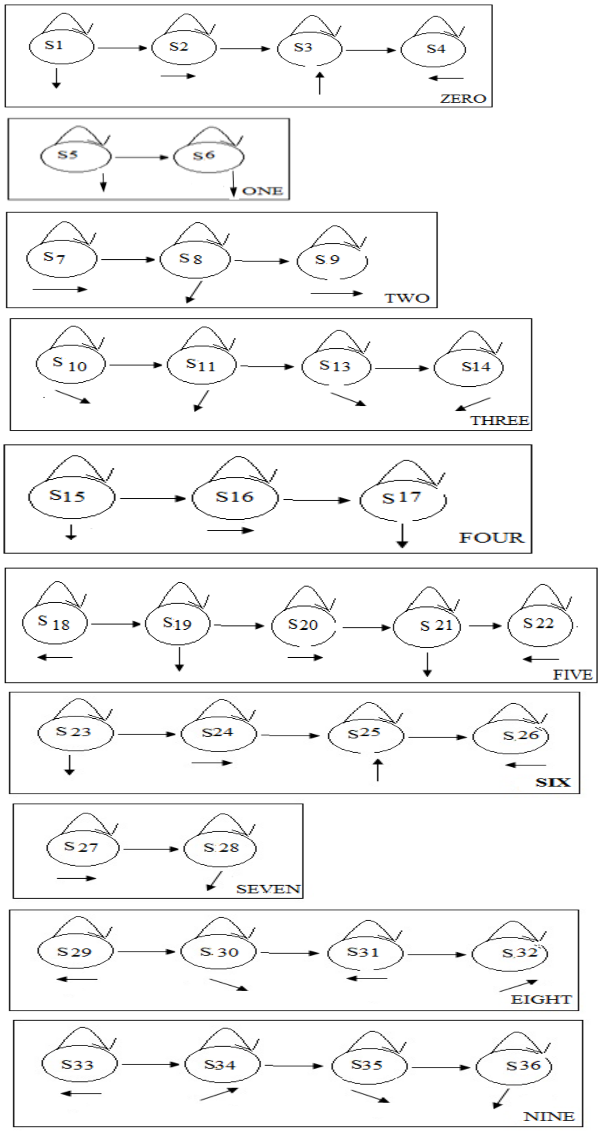



**Figure 4.8:** HMMs for digits 0 to 9.

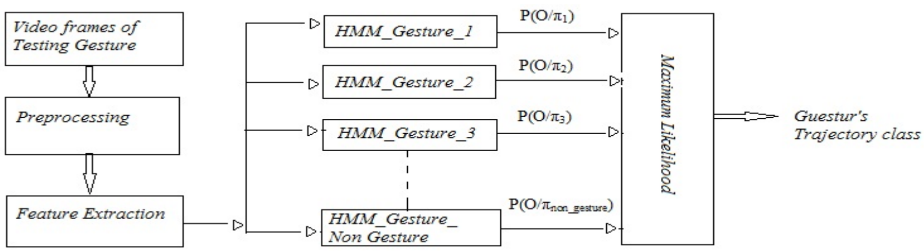



**Figure 4.9**: HMM for gesture trajectory classification

## 4.6 Experimental Results and Performance Evaluations

This chapter concentrates mainly on the motion trajectory analysis for dynamic hand gestures, and the suitability of the two feature vectors $f(\theta_1)$ and $f(\theta_2)$, for this purpose. The performance of the approach was evaluated on a video database containing numbers 0 to 9 represented (drawn in air) by hand motion. Experimental results are given in Table 4.2 and 4.3.

The results presented in table 4.2 indicates a high level of accuracy for numbers 1,4 and 7 which do not have much directional changes compared to the other numbers. At the same time, Table 4.3 gives good accuracy rate for numbers 2, 3, 8 and 0 which have frequent directional change. The results indicates that $f(\theta_1)$ is good for motion trajectories having minimum direction changes and f($\theta_2$) is good for motion trajectories have frequent directional changes. Since dynamic hand gestures in ISL exhibit both straight line motion and directional variations, a combination of f ($\theta_1$) and f ($\theta_2$) is needed for the recognition of motion trajectory as shown in Fig 4.10.Accuracy of the approach combining f($\theta_1$) and f($\theta_2$) is given in Table 4.4. An average accuracy of 93.19 % was obtained for the system.

For the complete recognition of a dynamic hand gesture, two steps are needed (Fig 4.2).

1. Identification of the hand shape
2. Identification of the motion trajectory.

Performance evaluation of the system is presented in Table 4.5 and graphical representation is shown in Fig 4.11. An accuracy of 91.2% was obtained for the complete recognition of the dynamic hand gestures. The gestures ‘GONE’ and ‘COME’ are the most complicated gestures and provided recall/sensitivity values equal to 0.6. In addition, the values were higher than 0.79 for all the statistical measures indicating better level of performance and suitability of oriented features in hand motion analysis for dynamic hand gesture recognition.

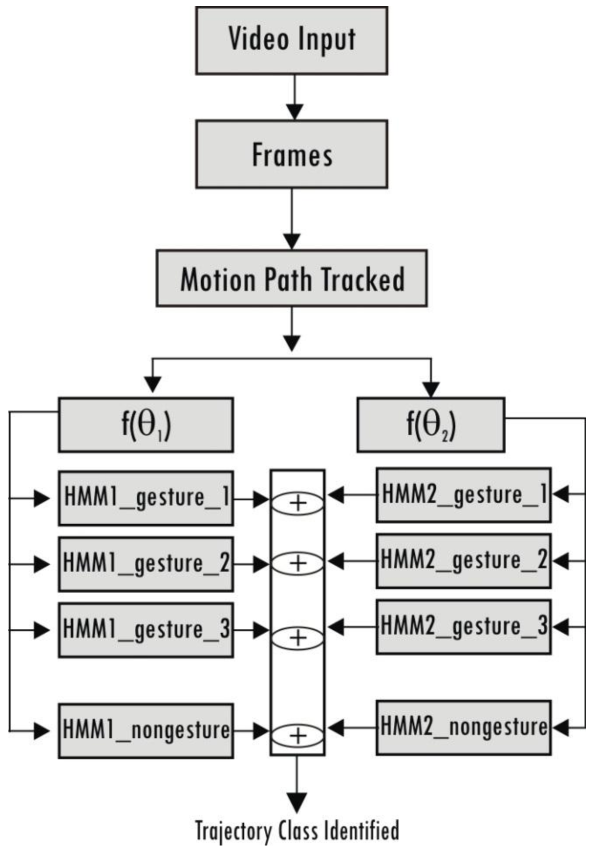



**Figure 4.10:** Combined f ($\theta_1$) and f ($\theta_2$) based HMM for motion trajectory classification.

**Table 4.2:** Confusion Matrix for the classifier based on f($\theta_1$).

| Actual Class Label | Identified Class | | | | | | | | | | | Accuracy rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Non_gesture set | |
| **0** | **24** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 88 |
| **1** | 0 | **26** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | **96.3** |
| **2** | 0 | 0 | **23** | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 85.2 |
| **3** | 0 | 0 | 2 | **23** | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 85.2 |
| **4** | 0 | 0 | 0 | 0 | **26** | 0 | 0 | 0 | 0 | 1 | 0 | **96.3** |
| **5** | 0 | 0 | 0 | 0 | 0 | **24** | 0 | 0 | 0 | 0 | 1 | 88 |
| **6** | 2 | 0 | 0 | 0 | 0 | 0 | **25** | 0 | 0 | 0 | 0 | 92.6 |
| **7** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | **26** | 0 | 0 | 0 | **96.3** |
| **8** | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | **22** | 0 | 1 | 81.5 |
| **9** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | **25** | 0 | 92.6 |

**Table 4.3:** Confusion Matrix for classifier based on f ($\theta_2$).

| Actual Class Label | Identified Class | | | | | | | | | | | Accuracy rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Non_gesture set | |
| **0** | **25** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 92.2 |
| **1** | 0 | **24** | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 88 |
| **2** | 0 | 0 | **26** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **96.3** |
| **3** | 0 | 0 | 1 | **26** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **96.3** |
| **4** | 0 | 1 | 0 | 0 | **24** | 0 | 0 | 2 | 0 | 0 | 0 | 88 |
| **5** | 1 | 0 | 0 | 0 | 0 | **25** | 0 | 0 | 0 | 0 | 1 | 92.2 |
| **6** | 1 | 0 | 0 | 1 | 0 | 0 | **25** | 0 | 0 | 0 | 0 | 92.2 |
| **7** | 0 | 1 | 1 | 0 | 1 | 0 | 0 | **24** | 0 | 0 | 0 | 88 |
| **8** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **26** | 0 | 1 | **96.3** |
| **9** | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | **25** | 0 | 92.2 |

**Table 4.4:** Confusion matrix for the classifier based on f ($\theta_1$) and f ($\theta_2$) for ISL signs.

| | Identified Class | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Actual Class Label | M | N | Z | J | Going | Coming | Happy | Sad | Gone | Come | Non_gesture set | Acc Rate |
| **M** | **25** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 92.6 |
| **N** | 0 | **25** | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 92.6 |
| **Z** | 0 | 0 | **26** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 96.2 |
| **J** | 0 | 0 | 0 | **26** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 96.2 |
| **Going** | 0 | 0 | 0 | 0 | **26** | 0 | 0 | 0 | 0 | 0 | 1 | 96.2 |
| **Coming** | 0 | 0 | 0 | 0 | 0 | **26** | 0 | 0 | 0 | 0 | 1 | 96.2 |
| **Happy** | 0 | 0 | 0 | 0 | 0 | 0 | **26** | 0 | 0 | 0 | 1 | 96.2 |
| **Sad** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **26** | 1 | 0 | 0 | 96.2 |
| **Gone** | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | **22** | 0 | 3 | 81.5 |
| **Come** | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **24** | 0 | 88 |

**Table 4.4:** Performance measures of Dynamic Hand Gesture Recognition System for ISL

| Gestures | Precision | Recall/Sensitivity | Specificity | F_measure | Accuracy |
|---|---|---|---|---|---|
| M | 0.71 | 0.63 | 0.91 | 0.66 | 0.87 |
| N | 0.75 | 0.69 | 0.96 | 0.71 | 0.86 |
| Z | 1 | 0.8 | 1 | 0.88 | 0.96 |
| J | 0.79 | 1 | 0.92 | 0.88 | 0.93 |
| Going | 0.7 | 0.82 | 0.91 | 0.75 | 0.94 |
| Coming | 0.7 | 1 | 0.89 | 0.82 | 0.96 |
| Happy | 1 | 0.8 | 1 | 0.88 | 0.96 |
| Sad | 1 | 1 | 1 | 1 | 1 |
| Gone | 0.68 | 0.6 | 0.91 | 0.63 | 0.81 |
| Come | 1 | 0.6 | 1 | 0.75 | 0.83 |

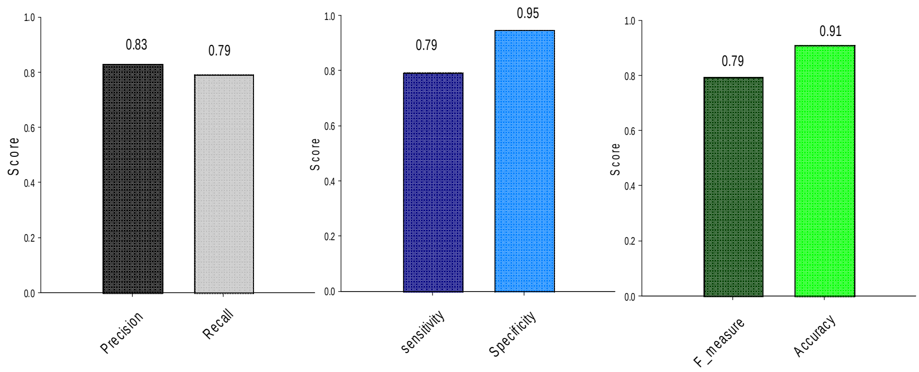



**Figure 4.11:** Average values of performance measures of dynamic hand gesture recognition system

## 4.7 Summary of the Chapter

Motion trajectory identification process is described in this chapter. It has been found that combined directional and orientation feature vectors are needed for better motion trajectory recognition.

……….☙❧……….

Chapter 5

# FACIAL EXPRESSION RECOGNITION USING GABOR FILTER AND DISPLACEMENT MEASURE VECTOR

Contents



*Certain hand gestures need facial expression changes also for conveying their complete meaning. This chapter describes an approach to recognize isolated facial expression change in simple ISL sentences using Gabor filter.*

## 5.1 Introduction

Facial expressions convey important information in sign language. For example, “where are you going?” in ISL is represented by two hand gestures (for ‘YOU’ and for ‘GOING’) and facial expression for ‘WHERE’ [88]. In addition, the emotions such as anger, happiness, surprise and sadness are conveyed by facial expression in ISL [89]. Without facial expression, signer cannot convey the complete intended meaning of the sentence. In these situations, one can notice transition of expressions from neutral stage to the peak expressive stage[88].

Accurately detecting the changes due to facial expression, its representation and classification are the major tasks involved. There are many challenges in the selection of facial expression features due to the dynamic nature of the signals that transmit information over time. Even the best classifiers fail to achieve accurate recognition rate due to inadequate features [90]. In facial expression analysis, there are mainly three types of approaches. The holistic approach based on appearance, the geometric feature based approach, and the hybrid approach based on appearance and geometry [91].

In the Geometric method, the location of key facial components such as mouth, eyes, eyebrow and nose are being tracked and any variation due to expression on these parts are being targeted [90]. Subsequently, the feature vector transmits the extracted facial components at these key geometric regions on face [92]. This analysis has wide applicability in exploiting facial representation.

The appearance based method, on the other hand, focuses on the whole face or specific region on face to frame the feature vector [90 ,91]. The feature vector targets the textural changes on face. Well known Holistic approaches are based on Principal Component Analysis [93]. Linear Discriminate Analysis (LDA) [94], Independent Component Analysis (ICA) [95] and Gabor wavelet analysis [96] which were applied to either the whole face or specific face regions can be used to extract the facial expression changes [90].

Facial expressions consistently lead to changes in skin textures by forming wrinkles and furrows (Fig 5.1).Gabor wavelet is well-known for capturing subtle textural changes on surface [90]. Also this Gabor wavelet representation has been successfully adopted in facial expression analysis [90]. In this chapter, we adopt, appearance based feature expression using Gabor Wavelet. Gabor wavelet parameters with Euclidian distance measure and Multi class SVM classifier were used to identify facial expressions in ISL.

In section 5.2 framework of the system is given. Section 5.3 gives an overview of features. Section 5.4 outlines the implementation. Section 5.5 analyses the result and the chapter summery is given in Section 5.6.

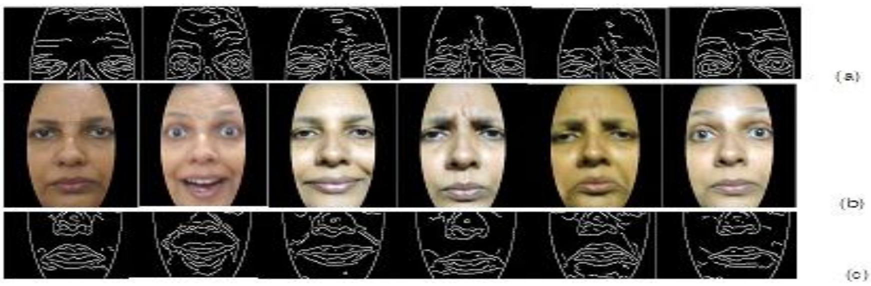



**Figure 5.1:** Row (b) shows the expressions neutral, surprise, happy, angry, sad and WH grammatical Marker (question). Rows (a) and (c) shows the wrinkle changes in the upper and lower area of the face during signing.

## 5.2 Recognition Framework

Facial expressions in ISL are described using facial feature changes. A simplified description of six main expressional changes considered in this chapter are represented in Fig 5.1.Our recognition system uses the changes that appear in the upper and lower face areas of the signer to classify the expressions during the information exchange. The semantic diagram of facial expression recognition system is given in Fig 5.2.

### 5.2.1 Image Acquisition

Videos of isolated expression sequences, representing short ISL sentences were recorded. In all the videos, signers' facial expression starts from neutral expression state and progresses into apex or peak point of expression. The length of the sequences varies depending on the facial expression and the signer/subject. The capturing is done using camera with frontal view of the person.

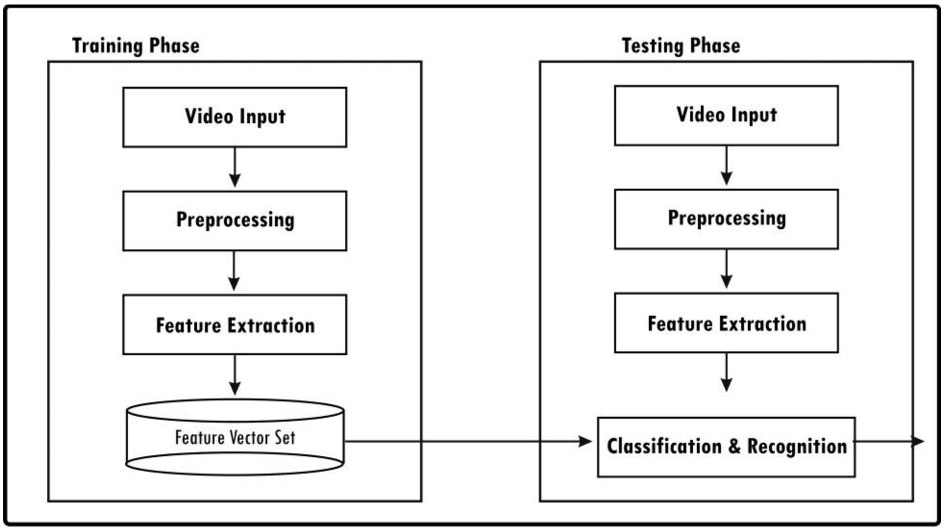



**Figure 5.2:** Facial Expression recognition system.

### 5.2.2 Pre-processing

In this phase, face region is extracted from the video frames. The two sub tasks involved in this operation are (1) detection of the face from the frames, (2) the extraction of frames with apex expressions (i.e. first and last frames). The cropped face area alone is used in recognition.

#### 5.2.2.1 Frame Extraction

In a video, there are two faces of interest. The neutral face and the apex face where maximum variation occurs. The most relevant information is present in the final face or apex face and is crucial in facial expression recognition [91]. Following this concept, the first frame representing neutral expression and last frame with peak expression are extracted for this study.

#### 5.2.2.2 Face Detection and Extraction

Aim of the module is to extract the face area from the first and apex frames. Viola and Jones method based on Haar-like features and the AdaBoost learning algorithm [92]is used to detect the face region from the frames. The face region is then cropped and masked into an elliptical shape as shown in Fig 5.3, to get the exact region of facial expression. This cropped face is then partitioned horizontally along the elliptical centre into upper and lower face regions. The RGB segmented face areas are then converted into a normalized gray scale image.

**Figure 5.3:** Steps in facial region extraction

## 5.3 Facial Expression Feature Extraction

The Fig5.1 clearly indicates the importance of wrinkles on facial expression recognition. This motivated us to apply the well-known textural analysis feature extraction method to the recognition process. Gabor wavelet method, which is well known for textural analysis and facial feature recognition, was chosen in our recognition system to represent the facial expressions. Facial expressions were described with the help of Gabor wavelet parameters. These feature parameters along with Euclidean distance measure were used to represent the facial expression changes in an image sequence.

### 5.3.1Gabor Wavelet Feature Representation

Gabor features were calculated by convolution of input image with Gabor filter bank [93,94]. Gabor filter works as a band pass filter for the local spatial frequency distribution and thereby achieve an optimal resolution in both spatial and frequency domain. The 2D Gabor filter $\psi(x,y,f,\theta)$ can be represented as a complex sinusoidal signal, modulated by a Gaussian kernel function as in Eq (5.1).

$$\psi(x,y,f,\theta) = [1/2\pi\sigma^2]\ [exp\{-(x_1^2 + y_1^2) / 2\sigma^2\}]\ [exp\ (2\pi f x_1)] \qquad (5.1)$$

where $x_1 = x \cos\theta + y \sin\theta$

$y_1 = - x \sin\theta + y \cos\theta$

σ is the standard deviation of Gaussian envelop along the x, y dimension, f is the central frequency of the sinusoidal plane wave, θ is the orientation of Gabor filter.

Feature extraction procedure can then be written as the convolution of gray scale facial expression image $I(x,y)$, with the Gabor filter $\psi(x,y,f,\theta)$ as in Eq (5.2).

$$G_{(u,v)}(x,y) = I(x,y) * \psi(x,y,f,\theta) \qquad (5.2)$$

In Eq (5.2), $G_{(u,v)}(x,y)$ represents the complex convolution output which can be decomposed into real and imaginary part as follows:

$E_{(u,v)}(x,y) = Re[G_{(u,v)}(x,y)]$ and $O_{(u,v)}(x,y) = I_m[G_{(u,v)}(x,y)]$.

Based on these results, both phase response and the magnitude response of the filter can be computed. In our work Gabor feature representation is based only on the magnitude response of the Gabor filter and phase response is neglected. Small spatial displacement causes significant variation in phase value. Due to this variation, the two Gabor features could not be directly compared. Magnitude response $A_{(u,v)}(x,y)$ of the filter can be computed as in Eq (5.3).

$$A_{(u,v)}(x,y) = \sqrt{E^2 + O^2} \quad \text{where } E = E_{(u,v)}(x,y) \text{ and } O = O_{(u,v)}(x,y) \qquad (5.3)$$

A Gabor filter bank with 5 frequencies and 8 orientations was used to extract Gabor features in our work as shown in Fig 5.5. Down sampling was done on all magnitude response, which were then normalized and concatenated into Gabor Feature Vector.


**Figure 5.4:** Gabor wavelet feature extraction sequence

### 5.3.2 Distance Measure on Gabor Feature Vector

Gabor wavelet feature vector representing magnitude information in 2D real matrix form was converted to one dimensional matrix without any loss in information. For a facial expression corresponding to a frame, feature vectors were generated for upper and lower face regions which were partitioned horizontally along the elliptical centre. In our work, we concentrated only on first and last frames in the video corresponding to isolated facial expressions. Six feature vectors were extracted from a video where two feature vectors representing the upper part of the faces in the first and last frame, and two vectors represent the lower parts of the face in the first and last frames and last two vectors represent the whole face regions. Later, the Euclidean distances between the corresponding vectors were found. In addition the percentage of change that occurred in the upper and lower face regions with respect to the total change between the neutral and apex frames were calculated. All these four measures formed the feature descriptors for the facial expression changes.

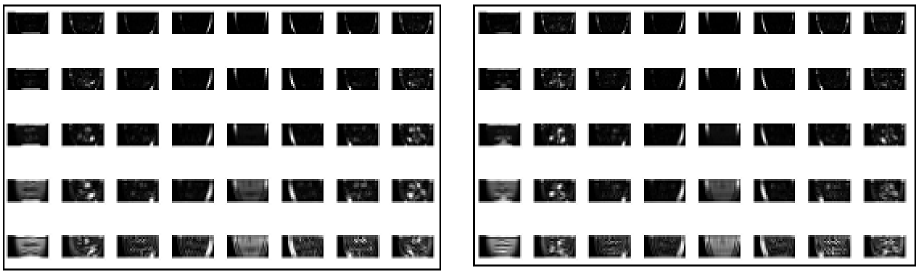


**Figure 5.5:** Gabor magnitude representations for the first and last frame's lower part of the face

## Dataset

Ten video frames of isolated facial expression for happiness, sadness, angry, surprise and question were collected from five persons at different time and location. The testing dataset was taken from different signers. All the extracted frames for training and testing were converted into normalized gray scale image of resolution 320 x 240 before processing.

## 5.4 Implementation

Facial expression recognition system consists of training and testing phase as shown in Fig 5.2. Pre-processing on a video input is done as described in section5.2.2. Feature vectors set generated in training phase was used by the classifier to recognize the facial expression gestures.

## 5.5 Experimental Results and Performance Evaluations

The experiment was conducted on videos exhibiting isolated facial expressions. Six common facial expression types were tested with Gabor wavelet methodology. Gabor wavelet parameters from the partitioned face areas, Euclidian distance measure and Multi class SVM classifier with RBF kernel function were used in this recognition system. Confusion matrix for recognizing facial expression changes is presented (Table 5.1). Performance measures of facial expression recognition system are given in Table 5.2 and graphical representation of them in Fig 5.6.The overall accuracy of 92.12 % was obtained. The average values of statistical measures were quite high and more than 0.81 (Fig 5.6). Experiment was also conducted on Cohn-Kanade facial expression dataset and got an average recognition rate of 94.28%. This clearly indicates the suitability of this method for facial expression recognition

**Table 5.1:** Confusion matrix for the classifier

| Expressions | HAPPY | ANGRY | SAD | SURPRISE | WH_expression | NEUTRAL |
|---|---|---|---|---|---|---|
| HAPPY | **0.9** | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| ANGRY | 0.0 | **0.9** | 0.0 | 0.0 | 0.1 | 0.0 |
| SAD | 0.1 | 0.0 | **0.9** | 0.0 | 0.0 | 0.0 |
| SURPRISE | 0.0 | 0.0 | 0.0 | **0.9** | 0.1 | 0.0 |
| WH_expression | 0.0 | 0.1 | 0.0 | 0.1 | **0.8** | 0.0 |
| NEUTRAL | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **1** |

**Table 5.2:** Performance measures of facial expression recognition system

| Expressional Changes | Precision | Recall/Sensitivity | Specificity | F_measure | Accuracy |
|---|---|---|---|---|---|
| Happy | 0.8 | 0.8 | 0.96 | 0.8 | 0.93 |
| Angry | 0.8 | 0.8 | 0.96 | 0.8 | 0.93 |
| Sad | 0.67 | 0.8 | 0.92 | 0.73 | 0.9 |
| WH_expression | 0.67 | 0.67 | 0.91 | 0.67 | 0.83 |
| Surprise | 1 | 0.8 | 1 | 0.89 | 0.93 |
| Neutral | 1 | 1 | 1 | 1 | 1 |

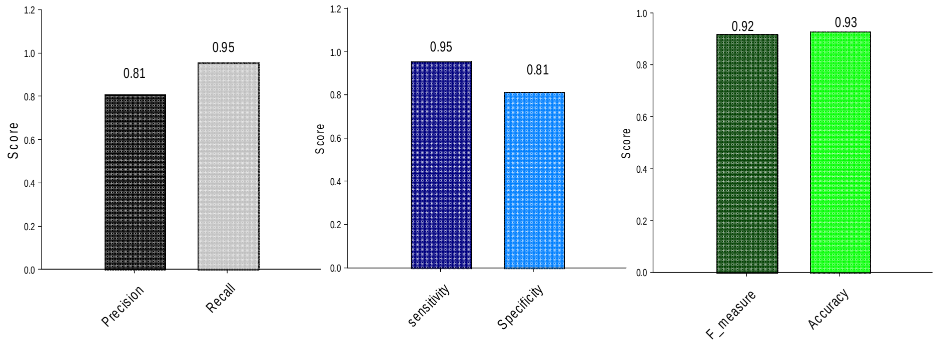



**Figure 5.6:** Comparison of average values of performance measures of facial expression recognition system.

## 5.6 Summary of the Chapter

The Gabor wavelet parameters with Euclidian distance measure and Multi class SVM classifier were used to identify facial expressions in ISL. The results indicated an overall accuracy of 92.12 % for the proposed system. The performance analysis revealed that the method followed in this study is highly promising and this could be used for the further up gradation of facial expression recognition systems in ISL.

*Chapter* **6**

# COMPUTATIONAL FRAMEWORK FOR INDIAN SIGN LANGUAGE SENTENCE RECOGNITION

*Contents*



*This chapter presents a system for the recognition of simple ISL sentences. The recognition schemes described in chapter 3, 4 and 5 are incorporated into a single framework for recognizing simple ISL sentences. This chapter also takes into account the grammar formalism for ISL sentences.*

## 6.1 Introduction

Sign language is the primary means of communication for the hearing and speaking impaired people. For communication, sign language uses different channels of signs as opposed to sound patterns. Importantly, sign languages evolved like any other spoken language and have its own grammar and syntax. All over the world, attempts have been made to integrate technological advancement and sign language linguistic advancements to provide better communication aids to the speaking and hearing impaired people.

In American Sign Language (ASL), Signing Exact English (SEE) modelling and recognition was done by integrating hand shape, orientation and location and movement [100]. Similarly, in another work, hand shape and facial expression are integrated based on LBP features [101]. However, these works cannot be replicated in ISL as every sign language differs and the rules and signs to represent sentences can vary. In this chapter a grammar formalism for ISL sentences was proposed. Sentences commonly used at lower primary school level were used to create the sentence and word database.

The chapter is organized as follows: Section 6.2 describes grammar formalism for ISL. Recognition framework for compound words and sentences in ISL in Section 6.3 and Section 6.4 respectively. Section 6.5 deals with performance analysis. The chapter concludes in Section 6.6

## 6.2 Grammar Formalism for ISL

Grammar formalism is a framework for explaining the basic structure of a language. For every language, there is syntax to represent sentences. Each sign language around the world has its own grammar. ISL has been reported to be a verb-ending language, i.e. Subject Object Verb (SOV) pattern. Information exchange in ISL or any other sign language is usually done through simple sentences. Although facial expression is very well described in ASL, research gaps do exist in ISL in this aspect [100].

A set of ISL sentences of primary school level used in this work is given in Table 6.1. Some of the characteristics exhibited by ISL sentences are:

- Sentences follow S-O-V pattern.
- Adjectives, verbs and WH words come at the end of the sentences.
- Interrogative sentences and sentences with adjectives are expressed with facial expressions

An Augmented Transition Network (ATN) can be used to represent the basic structure of simple ISL sentences as shown in Fig 6.1. ATN is a finite state machine [102].

**Table 6.1:** List of simple English sentences and there equivalents in ISL

| Sentences in English | Equivalent ISL sentences |
|---|---|
| Cat caught rat. | [Cat Rat Catch] |
| I like apple. | [I, Apple, Like] |
| I have four brothers. | [I, Brother ,Four] |
| My mother is coming. | [My, Mother, Coming] |
| My friend has gone. | [My, Friend, Gone] |
| I like my sister | [I, Sister, Like] |
| What is your name? | [You, Name, What] |
| I am hungry. | [I, Hungry] |
| Where are you going? | [You, Going, Where] |
| I like my friend. | [I, Friend, Like] |
| Who is your friend: | [You, Friend , Who] |
| I am going home. | [I, House, Going] |
| What food do you like? | [Food, You, Like ,What] |
| I am happy today. | [I, Today, Happy] |
| No school on Sunday. | [School, Sunday, No] |
| Grandfather has gone home. | [Grandfather, House, Gone] |
| Today is Monday. | [Today, Monday] |
| Teacher is coming. | [Teacher, Coming] |
| I like blue color. | [I, Blue, like] |
| Sister likes reading. | [Sister, Read, Like] |
| Father and mother are coming together. | [Father, Mother, Together, Coming] |
| You are a good child. | [You , Child, Good] |
| When you come? | [You, Come, When] |
| What are you thinking? | [You, Think, What] |
| Ball is red color. | [Ball, Red] |

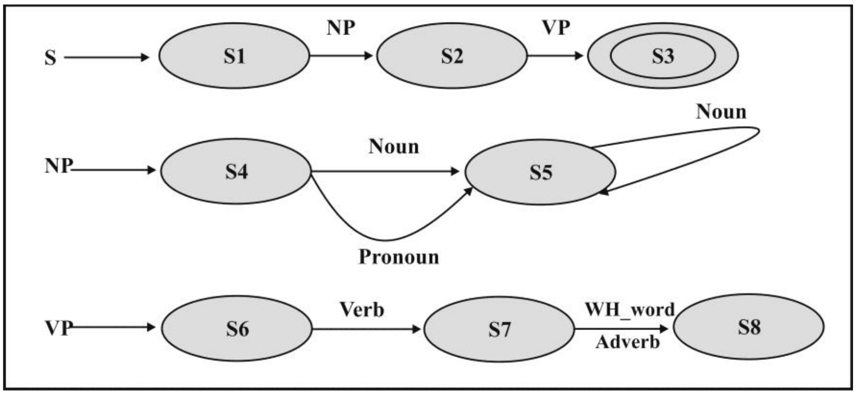



**Figure 6.1**: Augmented Transition Network for simple ISL sentences.

## 6.3 Recognition Framework for Compound words in ISL

Words that are represented by more than one sign are referred as compound words. In ISL, signs in compound words are static hand postures. Fig 6.2 shows the representation of APPLE, which is a compound word that is signed by two signs. In ISL, compound words are represented by two signs.

**Figure 6.2:** APPLE

Recognition system for compound words was developed by the scheme described in chapter 3. Gesture spotting and recognition are the two main tasks involved in a compound word recognition. The frame work of the system is given in Fig 6.3.

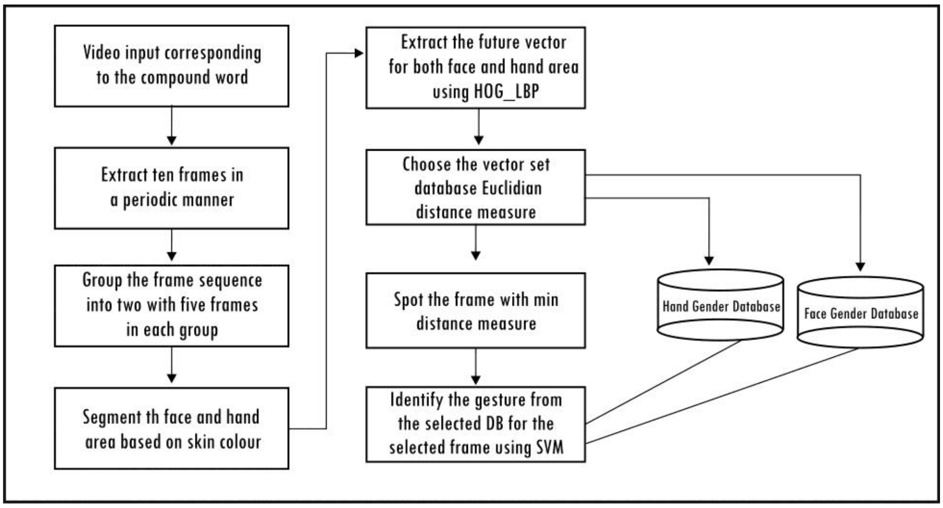



**Figure 6.3:** Recognition framework for compound words.

### 6.3.1 Implementation of Compound word Recognition

A set of signs that may form part of compound words were collected and stored as training data. Later this dataset was used in recognition phase.

**Recognition phase**

Steps involved in the recognition of compound sign words are summarized as follows:

1. Input video corresponding to the compound word
2. Convert the video input to frame sequence.
3. Select ten frames in a periodic manner based on the total frames extracted.
4. Group the sequence into two, with five frames of each.
5. Identify face and hand area based on skin color. Upper largest connected component is face and lower largest component is hand,(Fig 6.4)
6. Extract feature vectors for both hand and face in each frame.
7. Select the corresponding database for the two groups using minimum Euclidian distance measure.
8. Spot the frame in each group that gives minimum distance measure, from hand gesture database/face gesture database.

9. Identify the correct gesture using SVM classifier from the corresponding selected database
10. Compound word is identified as a sequence of the two gesture classes recognized in the previous step.

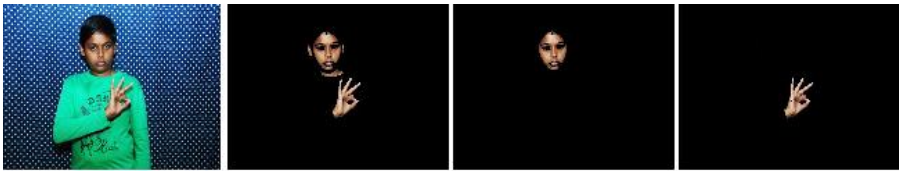


**Figure 6.4:** Face and hand extracted from a frame.

## 6.4 Recognition Framework for simple ISL sentence

ISL sentence is a combination of all the three channels of signs. One or more noun groups followed by a verb group are considered as a simple sentence. The schematic diagram for gesture spotting in a simple ISL sentence is shown in Fig 6.5.

### 6.4.1 Implementation of simple ISL sentence recognition

A set of gestures from all classes are used in this study (static gesture, dynamic gesture and facial expression). Database of feature vector set is generated separately for all classes of gestures. In addition, one separate database for hand-on-face gestures are created (total four database).All the feature vectors are generated based on the decisive feature descriptors discussed in the previous chapters.

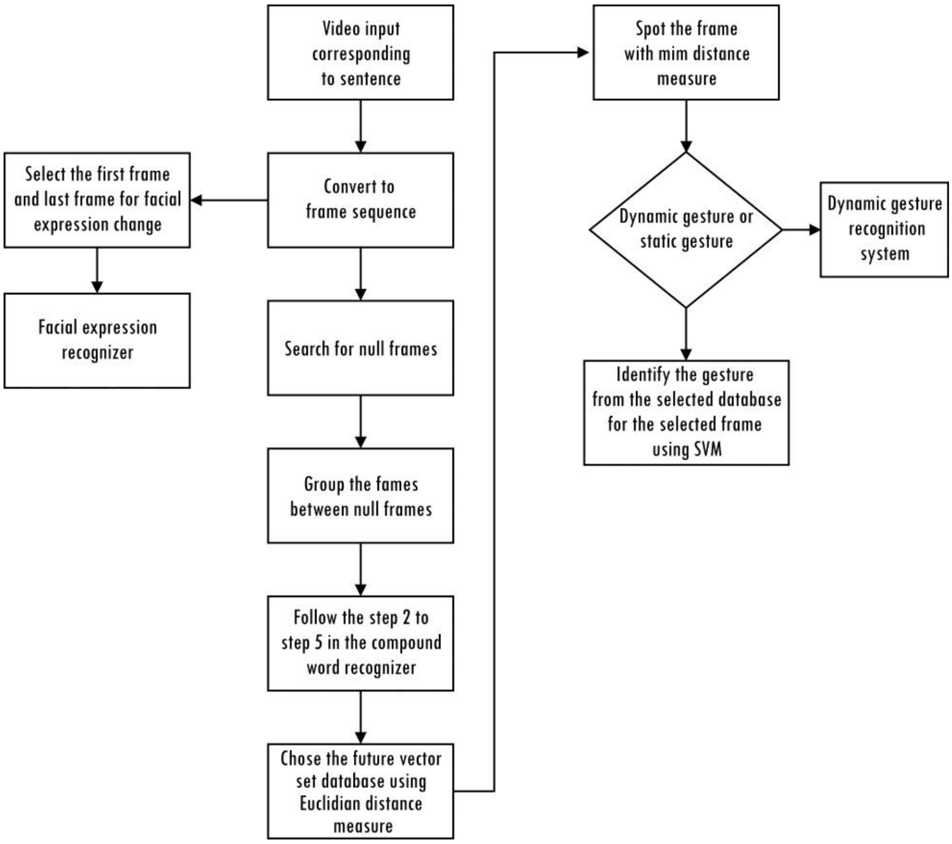



**Figure 6.5:** Schematic diagram for gesture spotting in simple ISL sentence

**Recognition phase**

Steps for simple ISL sentence recognition are summarized below:

1. Input video corresponding to the sentence
2. Convert the video input into frame sequences. Pass the first and last frame into facial expression recognizer. Pass total frames for the hand gesture recognition process.
3. Segment the hand and face area based on skin color
4. Search for the first null frame, which contain only face with no gestures
5. Group the frames based on null frame.
6. Consider ten frames in a periodic manner based on the total frames in the group
7. Group the sequences into two with five frames each.
8. Extract the feature vectors corresponding to face (largest upper connected components) and hand (largest lower connected component) for each frame.
9. Hand feature vectors from the first group are compared with static hand gesture feature set and dynamic feature set. If it gives

minimum distance for dynamic feature set then total frame group consist of ten frames will pass to dynamic gesture recognizer, otherwise it will go to next channel for comparison.

10. Feature vectors corresponding to face and hand of each frame from the two groups are compared with static hand gesture and overlaid hand on face feature sets.
11. Select the dataset based on minimum Euclidian distance measure. Also select frames and the area (face or hand) in each group that gives minimum distance measure.
12. The two gestures spotted from two groups are identified using SVM classifier with the corresponding dataset.
13. If two gestures are identified, then it would be considered as compound gesture, otherwise treat it as simple gesture.
14. Go to step 5 for selecting the next group, till the last frame is identified.
15. Order the identified gesture class to generate the sentence along with the facial expression changes identified.

***Null Frame identification***

Null frame selection is as follows:

1. Look for frames with only upper component (face)
2. Generate feature vector for the corresponding frames.
3. If the Euclidian similarity score is greater than 15 with the overlaid feature vector set, then select the frame as null frame, otherwise continue for searching.

## 6.5 Performance Analysis

The work was evaluated by a set of ISL sentences of primary school level. The ISL sentences followed Subject-Object-Verb pattern. The dataset contained about 75 signs as given in Appendix A. Feature vectors representing signs were kept in four datasets. Testing was done on video dataset which is a collection of compound words and simple ISL sentences. Tables 6.2 summarize the result of recognition rate of compound words. Some of the ISL sentences used for the evaluation of the system and the success obtained is tabulated in Table 6.3. Total 23 words were identified from 28 words which were performed as part of expressing sentences. ISL simple sentence recognition system showed a gesture recognition rate of 82.14% in this experiment. Interrogative sentences exhibited a recognition rate of 86.7%.

**Table 6.2:** Recognition rate of words

| Word Gestures | Recognition Rate |
|---|---|
| Human Relationship Words (compound words)<br>(Father, Mother, Sister, Brother, Grandfather, Grandmother) | 83.33% |
| Days of a week<br>(Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday) | 89 % |
| Common words<br>(Apple, Tomato, Teacher, Ball, Red ,Blue, Black, House, Like, Friend, Together etc) | 79.08% |

**Table 6.3:** Recognition of Simple Sentences

| Sentences in English | Equivalent ISL Sentences | Gestures Identified | Not Identified | Followed SOV Pattern |
|---|---|---|---|---|
| Cat caught rat | Cat Rat Catch | All got identified | | YES |
| Sister is coming today | Sister Today Coming | Today, Coming | Sister Partially identified. Second part of sister was not identified. | YES |
| I like apple | I Apple Like | I, Like | First part of apple was not identified | YES |
| Teacher is coming | Teacher Coming | All got identified | | YES |
| Where are you going? | You Going Where | Going, Where | You | YES |
| I am going home. | I Home Going | All got identified | | YES |
| I like grandfather. | I Grandfather like | I, Like | Null frame got identified between Grandfathers signs | NO |
| What is your name? | You Name What | Name , What | You | YES |
| Today is Monday. | Today Monday | All got identified | | YES |
| What food do you like? | Food Like What | All got identified | | YES |

## 6.6 Summary of the chapter

This chapter presents a computational framework for recognizing ISL sentences and compound words by integrating different channels of sign. The framework is implemented using the decisive feature descriptors identified in the earlier chapters. The performance analysis revealed that the methods followed in this study are highly promising and this could be used for the future upgradation of ISL recognition systems.

Chapter 7

# CONCLUSION AND FUTURE WORKS

Contents



## 7.1 Conclusions

This research work is about the development of a computational framework for Indian Sign Language recognition. ISL recognition mainly involves recognition of static hand gestures, dynamic hand gestures and facial expressions. All these three channels of ISL communication are associated with several inherent complexities.

The complexities addressed in this work include static hand gestures with resemblances, overlaid hand gestures, similar dynamic gestures with change in motion trajectory and isolated facial expression changes in sentences. For gestures with resemblances, a detailed analysis and experimentation with different types of feature descriptors was carried. The best performance was achieved by HOG feature descriptors for gestures with resemblance while HOG-LBP feature descriptors gave better results for overlaid gestures. Instead of a holistic or local approach, a semi global approach was found to be ideal for static hand gestures.

In this study, for dynamic gesture recognition a better approach was derived. It has been found that combined directional and orientation feature vectors are used for better motion trajectory recognition, while the existing conventional approaches used only orientation angle change,

This work also explored complexities in the recognition of facial expression changes in ISL sentences, an area where not much research has been done. Performance analysis revealed that feature descriptors using Gabor Wavelet and Euclidean measure were highly promising in enhancing gesture recognition accuracy. This is a promising result for the further upgradation of facial expression recognition system in ISL.

Statistical measures were used for analysing the performance of the system.

## 7.2 Future Works

The current study opened some future research directions that could improve ISL recognition.

- The combinations of approaches identified can be applied on a broad data base of gestures for improving the precision and accuracy of gesture recognition substantially.
- Simple ISL sentence recognition done in this work can be extended to compound ISL sentence recognition.
- This work can be extended for the recognition of multiple facial expressional changes in ISL sentences.
- Standardization of computational framework for the recognition of ISL which has got tremendous application in education and socio cultural areas of hard to hear and speak people.

# REFERENCES


1. Ong, Sylvie CW, and Suhas Ranganath. "Automatic sign language analysis: A survey and the future beyond lexical meaning." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.6 (2005): 873-891.

2. Mitra, Sushmita, and Tinku Acharya. "Gesture recognition: A survey." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 37.3 (2007): 311-324.

3. Sahoo, Ashok K., Gouri Sankar Mishra, and Kiran Kumar Ravulakollu. "Sign Language Recognition: State of the Art." *ARPN Journal of Engineering and Applied Sciences* 9.2 (2014): 116-134.

4. McGuire, R. Martin, et al. "Towards a one-way American sign language translator." *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*. IEEE, 2004.

5. Holden, Eun-Jung, Gareth Lee, and Robyn Owens. "Australian sign language recognition." *Machine Vision and Applications* 16.5 (2005): 312-320.

6. Starner, Thad, and Alex Pentland. "Real-time American sign language recognition from video using hidden markov models." *Motion-Based Recognition*. Springer Netherlands, 1997. 227-243.

7. Alon, Jonathan, et al. "A unified framework for gesture recognition and spatiotemporal gesture segmentation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.9 (2009): 1685-1699.

8. Athitsos, Vassilis, Haijing Wang, and Alexandra Stefan. "A database-based framework for gesture recognition." *Personal and Ubiquitous Computing* 14.6 (2010): 511-526.

9. Bhat, Nagaraj N., et al. "Hand gesture recognition using self organizing map for Human Computer Interaction." *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*. IEEE, 2013.

10. Oz, Cemil, and Ming C. Leu. "Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker." *Neurocomputing* 70.16 (2007): 2891-2901.

11. Derpanis, K.G., R.P. Wildes, and J.K. Tsotsos, *Definition and recovery of kinematic features for recognition of American sign language movements.* Image and Vision Computing, 2008. 26(12): p. 1650-1662.

12. Ding, L. and A.M. Martinez, *Modelling and recognition of the linguistic components in American Sign Language.* Image and Vision Computing, 2009. 27(12): p. 1826-1844.

13. Elmezain, Mahmoud, et al. "Spatio-temporal feature extraction-based hand gesture recognition for isolated American Sign Language and Arabic numbers." *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*. IEEE, 2009.

14. Ullah, Fahad. "American Sign Language recognition system for hearing impaired people using Cartesian Genetic Programming." *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*. IEEE, 2011.

15. Tangsuksant, Watcharin, Suchin Adhan, and Chuchart Pintavirooj. "American Sign Language recognition by using 3D geometric invariant feature and ANN classification." *Biomedical Engineering International Conference (BMEiCON), 2014 7th*. IEEE, 2014.

16. Gupta, Lalit, and Suwei Ma. "Gesture-based interaction and communication: automated classification of hand gesture contours." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 31.1 (2001): 114-120.

17. Kong, W. W., and Surendra Ranganath. "Sign language phoneme transcription with rule-based hand trajectory segmentation." *Journal of Signal Processing Systems* 59.2 (2010): 211-222.

18. Isaacs, Jason, and Simon Foo. "Hand pose estimation for American sign language recognition." *System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on*. IEEE, 2004.

19. Kim, Taehwan, Karen Livescu, and Greg Shakhnarovich. "American sign language fingerspelling recognition with phonological feature-based tandem models." *Spoken Language Technology Workshop (SLT), 2012 IEEE*. IEEE, 2012.

20. Kurakin, Alexey, Zhengyou Zhang, and Zicheng Liu. "A real time system for dynamic hand gesture recognition with a depth sensor." *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012.

21. Liang, Rung-Huei, and Ming Ouhyoung. "A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System." *Computer Graphics Forum*. Vol. 14. No. 3. Blackwell Science Ltd, 1995.

22. Liu, Jingjing, et al. "Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions." *Image and Vision Computing* 32.10 (2014): 671-681.

23. Rashid, Omer, Ayoub Al-Hamadi, and Bernd Michaelis. "A framework for the integration of gesture and posture recognition using HMM and SVM." *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*. Vol. 4. IEEE, 2009.

24. Rashid, Omer, Ayoub Al-Hamadi, and Bernd Michaelis. "Utilizing invariant descriptors for finger spelling american sign language using svm." *Advances in Visual Computing*. Springer Berlin Heidelberg, 2010. 253-263.

25. Nguyen, Tan Dat, and Surendra Ranganath. "Facial expressions in American sign language: Tracking and recognition." *Pattern Recognition* 45.5 (2012): 1877-1891.

26. Derpanis, Konstantinos G., Richard P. Wildes, and John K. Tsotsos. "Definition and recovery of kinematic features for recognition of American sign language movements." *Image and Vision Computing* 26.12 (2008): 1650-1662.

27. Al-Jarrah, Omar, and Faruq A. Al-Omari. "Improving gesture recognition in the Arabic sign language using texture analysis." *Applied Artificial Intelligence* 21.1 (2007): 11-33.

28. Tharwat, Alaa, et al. "SIFT-Based Arabic Sign Language Recognition System." *Afro-European Conference for Industrial Advancement*. Springer International Publishing, 2015.

29. Tubaiz, Noor Ali. *Sensor-based Continuous Arabic Sign Language Recognition*. Diss. American University of Sharjah, 2014.

30. Al-Rousan, M., KhaledAssaleh, and A. Tala'a. "Video-based signer-independent Arabic sign language recognition using hidden Markov models." *Applied Soft Computing* 9.3 (2009): 990-999.

31. Shanableh, Tamer, KhaledAssaleh, and M. Al-Rousan. "Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language." *Systems, Man,*

*and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 37.3 (2007): 641-650.

32. Tolba, M. F., Ahmed Samir, and Magdy Aboul-Ela. "Arabic sign language continuous sentences recognition using PCNN and graph matching." *Neural Computing and Applications* 23.3-4 (2013): 999-1010.

33. Elons, A. S., M. Ahmed, and H. Shedid. "Facial expressions recognition for arabic sign language translation." *Computer Engineering & Systems (ICCES), 2014 9th International Conference on*. IEEE, 2014.

34. Mohandes, Mohamed A. "Recognition of two-handed Arabic signs using the Cyber Glove." *Arabian Journal for Science and Engineering* 38.3 (2013): 669-677.

35. Mohandes, M., S. Aliyu, and M. Deriche. "Arabic sign language recognition using the leap motion controller." *Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on*. IEEE, 2014.

36. Mohandes, Mohamed, and Mohamed Deriche. "Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence." *Computing, Communications and IT Applications Conference (ComComAp), 2013*. IEEE, 2013.

37. Aujeszky, Tamás, and Mohamad Eid. "A gesture recognition architecture for Arabic sign language communication system." *Multimedia Tools and Applications* (2015): 1-19.

38. Aly, Saleh, and Safaa Mohammed. "Arabic Sign Language Recognition Using Spatio-Temporal Local Binary Patterns and Support Vector Machine." *Advanced Machine Learning Technologies and Applications*. Springer International Publishing, 2014. 36-45.

39. Aly, Sherin. "Appearance-based Arabic Sign Language recognition using Hidden Markov Models." *Engineering and Technology (ICET), 2014 International Conference on*. IEEE, 2014.

40. Shanableh, Tamer, and Khaled Assaleh. "Video-based feature extraction techniques for isolated Arabic sign language recognition." *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*. IEEE, 2007.

41. Tolba, M. F., Ahmed Samir, and Magdy Abul-Ela. "A proposed graph matching technique for Arabic sign language continuous sentences recognition." *Informatics and Systems (INFOS), 2012 8th International Conference on*. IEEE, 2012.

42. Fang, G., W. Gao, and D. Zhao, Large-vocabulary continuous sign language recognition based on transition-movement models. Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 2007. 37(1): p. 1-9.

43. Fang, G.L. and W. Gao, A SOFM/HMM system for person-independent isolated sign language recognition. Human-Computer Interaction - Interact'01, ed. M. Hirose2001. 731-732.

44. Fang, G.L., W. Gao, and S. Ieee Computer Society; Ieee Computer, A SRN/HMM system for signer-independent continuous sign language recognition. Fifth Ieee International Conference on Automatic Face and Gesture Recognition, Proceedings2002. 312-317.

45. Fang, G.L., et al., Signer-independent sign language recognition based on SOFM/HMM. Ieee Iccv Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, Proceedings2001. 90-95.

46. Fang, G.L., et al., A novel approach to automatically extracting basic units from Chinese sign language, in Proceedings of the 17th International Conference on Pattern Recognition, Vol 4, J. Kittler, M. Petrou, and M. Nixon, Editors. 2004. p. 454-457.

47. Quan, Y. and Ieee, Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling. Iciea 2010: Proceedings of the 5th Ieee Conference on Industrial Electronics and Applications, Vol 32010. 385-390.

48. Wang, C.L., et al., An approach based on phonemes to large vocabulary Chinese sign language recognition. Fifth Ieee International Conference on Automatic Face and Gesture Recognition, Proceedings2002. 411-416.

49. Wang, C.L., W. Gao, and Z.G. Xuan, A real-time large vocabulary continuous recognition system for Chinese Sign Language, in Advances in Multimedia Information Processing - Pcm 2001, Proceedings, H.Y. Shum, M. Liao, and S.F. Chang, Editors. 2001. p. 150-157.

50. Zhou, Y., et al., Mahalanobis Distance Based Polynomial Segment Model For Chinese Sign Language Recognition. 2008 Ieee International Conference on Multimedia and Expo, Vols 1-42008. 317-320.

51. Zhou, Y., et al., Signer adaptation based on etyma for large vocabulary Chinese sign language recognition, in Advances in Multimedia Information Processing - Pcm 2007, H.H.S. Ip, et al., Editors. 2007. p. 458-461.

52. Nandy, Anup, et al. "Recognition of isolated Indian sign language gesture in real time." *Information Processing and Management*. Springer Berlin Heidelberg, 2010. 102-107.

53. Rekha, J., J. Bhattacharya, and S. Majumder. "Shape, texture and local movement hand gesture features for Indian sign language recognition." *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on*. IEEE, 2011.

54. Kishore, P. V. V., et al. "Video Audio Interface for Recognizing Gestures of Indian Sign." *International Journal of Image Processing (IJIP)* 5.4 (2011): 479.

55. Tewari, Deepika, and Sanjay Kumar Srivastava. "A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm." *International Journal of Engineering and Advanced Technology (IJEAT)* 2.2 (2012): 165-170.

56. Futane, Pravin R., and Rajiv V. Dharaskar. "Video gestures identification and recognition using Fourier descriptor and general fuzzy minmax neural network for subset of Indian sign language." *Hybrid Intelligent Systems (HIS), 2012 12th International Conference on*. IEEE, 2012.

57. Singha, Joyeeta, and Karen Das. "Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique." *arXiv preprint arXiv: 1303.0634* (2013).

58. Dour, Shweta, and J. M. Kundargi. "Article: Design of ANFIS System for Recognition of Single Hand and Two Hand Signs for Indian Sign Language}." *IJAIS Proceedings on International Conference and workshop on Advanced Computing* . 2013.

59. Geetha, M., and U. C. Manjusha. "A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation." *International Journal on Computer Science and Engineering* 4.3 (2012): 406.

60. Saraswat, Mukesh, and K. V. Arya. "Automatic Facial Expression Recognition in an Image Sequence of Non-manual Indian Sign Language Using Support Vector Machine." *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*. Springer India, 2012.

61. Geetha, M., et al. "A vision based dynamic gesture recognition of Indian Sign Language on Kinect based depth images." *Emerging Trends in Communication, Control, Signal*

*Processing & Computing Applications (C2SPCA), 2013 International Conference on*. IEEE, 2013.

62. Gupta, Prerna, Garima Joshi, and Maitreyee Dutta. "Comparative Analysis of Movement and Tracking Techniques for Indian Sign Language Recognition." *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on*. IEEE, 2015.

63. Karami, Ali, Bahman Zanj, and Azadeh Kiani Sarkaleh. "Persian sign language (PSL) recognition using wavelet transform and neural networks." *Expert Systems with Applications* 38.3 (2011): 2661-2667.

64. Teng, Xiaolong, et al. "A hand gesture recognition system based on local linear embedding." *Journal of Visual Languages & Computing* 16.5 (2005): 442-454.

65. Hrúz, M., J. Trojanová, and M. Železný. "Local Binary Pattern based features for sign language recognition." *Pattern Recognition and Image Analysis* 22.4 (2012): 519-526.

66. Von Agris, Ulrich, Moritz Knorr, and Karl-Friedrich Kraiss. "The significance of facial features for automatic sign language recognition." *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008.

67. Indian Sign Language Single Handed. http:// Indian sign language. org/isldictionary/alphabets/single-handed (accessed 10 Mar. 2013).

68. Collumeau, Jean-François, et al. "Hand gesture recognition using a dedicated geometric descriptor." *Image Processing Theory, Tools and Applications (IPTA), 2012 3rd International Conference on*. IEEE, 2012

69. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, 2005.

70. Ojala, Timo, MattiPietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." *Computer Vision-ECCV 2000*. Springer Berlin Heidelberg, 2000. 404-420.

71. Mäenpää, Topi. *The Local binary pattern approach to texture analysis: Extenxions and applications*. Oulunyliopisto, 2003.

72. Kovač, Jure, Peter Peer, and Franc Solina. *Human skin color clustering for face detection*. Vol. 2. IEEE, 2003.

73. Wang, Chieh-Chih, and Ko-Chih Wang. "Hand Posture recognition using Adaboost with SIFT for human robot interaction." *Recent progress in robotics: viable robotic service to human*. Springer Berlin Heidelberg, 2008. 317-329.

74. Yao, Yuan, and Yun Fu. "Real-time hand pose estimation from RGB-D sensor." *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 2012.

75. Yun, Liu, and Zhang Peng. "An automatic hand gesture recognition system based on Viola-Jones method and SVMs." *Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on*. Vol. 2. IEEE, 2009.

76. Dardas, Nasser H., and Emil M. Petriu. "Hand gesture detection and recognition using principal component analysis." *Computational Intelligence for Measurement Systems and Applications (CIMSA), 2011 IEEE International Conference on*. IEEE, 2011.

77. Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter," Annual Conference on Computer Graphics & Interactive Techniques. ACM Press, Addison-Wesley, Los Angeles, CA, USA (August 12–17), SIGGRAPH 2001 course pack edition.

78. Silanon, Kittasil, and Nikom Suvonvorn. "Hand motion analysis for Thai alphabet recognition using HMM." *International Journal of Information and Electronics Engineering* 1.1 (2011): 65-71.

79. Ramamoorthy, Aditya, et al. "Recognition of dynamic hand gestures." *Pattern Recognition* 36.9 (2003): 2069-2081.

80. Liu, Nianjun, et al. "Model structure selection & training algorithms for an HMM gesture recognition system." *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*. IEEE, 2004.

81. Elmezain, Mahmoud, Ayoub Al-Hamadi, and Bernd Michaelis. "Real-time capable system for hand gesture recognition using hidden markov models in stereo color image sequences." (2008).

82. Elmezain, Mahmoud, et al. "A hidden markov model-based continuous gesture recognition system for hand motion trajectory." *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008.

83. Kao, Chang-Yi, and Chin-Shyurng Fahn. "A human-machine interaction technique: hand gesture recognition based on hidden Markov models with trajectory of hand motion." *Procedia Engineering* 15 (2011): 3739-3743.

84. Li, Chao-Tang, and Wen-Hui Chen. "A Novel FPGA-based Hand Gesture Recognition System." *Journal of Convergence Information Technology* 7.9 (2012).

85. Wang, Xiaoyan, et al. "Hidden-markov-models-based dynamic hand gesture recognition." *Mathematical Problems in Engineering* 2012 (2012).

86. Ioannis "Gesture Recognition for Alphabet Characters from Fingertip Motion Trajectory Using LRB Hidden Markov Models"

87. Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989): 257-286.

88. M.W. Morgan. Topology of Indian Sign Language verbs from a comparative perspective An annual review of South Asian languages and linguistics,vol 222, pp. 103-131, 2009

89. B. Bridges and M. Metzger. Deaf tend your. Calliope press. Silver Spring, 1996, pp 27-38 .

90. Y.L. Tiam, T. Kanade and J.F. Cohn. Facial expression analysis, Springer New York, 2005,, pp. 247-275.

91. C. Shan, S. Gong and P.W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, vol. 27, pp. 803-816, 2009.

92. D. Das. Human's facil parts extraction to recognize facial expression. International journal on information theory, vol. 3, pp. 65-72. 2014.

93. M. Turk and A.P. Pentlanad. Face recognition using Eigen face. IEEE Conference on computer vision and pattern recognition, 1991.

94. P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman. Eigen faces versus Fisher faces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence, pp. 711-720, 1997.

95. Bartlett, M.S., J.R. Movellan and T.J. Sejnowski. Face recognition by independent component analysis. IEEE Transactions on Neural Networks, pp. 1450-1464, 2002.

96. M.J. Lyions, J. Budynek and S. Akametsu. Automatic classification of single facial image. IEEE Transactions on pattern analysis and machine intelligence, pp. 1357-1362, 1999.

97. J.M. Gold. Efficiency of dynamic and static facial expression. Journal of vision, pp. 1-12, 2013

98. S. Sindhumol, K. Anil, and B. Kannan, “Abnormality detection from multispectral brain MRI using multi resolution independent component analysis”, International Journal of Signal Processing, Image Processing and Pattern Recognition. 2013, pp177-190

99. S. S. K. Nair**,** N. V. Subba Reddy and K.S. Hareesha, **“**Exploiting heterogeneous features to improve *in silico* prediction of peptide status amyloidogenic or non-amyloidogenic”, *BMC Bioinformatics,* 2011,pp 2-9.

100. Bridges, Byron, and Melanie Metzger. "Deaf tend your. Non manual signals in ASL. Salem, Or.: Sign Enhancers." (1997).

101. www.iitg.ernet.in/isl/

102. An introduction on ATN in LISP by Paul Graham

# LIST OF PUBLICATIONS

## Papers in International Conferences:

1. Daleesha M Viswanathan, Sumam Mary Idicula, *'Hand Gesture Recognition Of English Alphabets Using Semi Global Descriptor'* *International Conference on Image and Signal* Processing, ICISP-2014,Bangalore, Elsevier India.
2. Daleesha M Viswanathan ,Sumam Mary Idicula, *'Recognition of Hand Gestures of English Alphabets using HOG Method'* IEEE 2014 International Conference on Data Science & Engineering (ICDSE),CUSAT
3. Daleesha M Viswanathan, Sumam Mary Idicula, "*Hand Motion Trajectory Analysis for Dynamic Hand Gestures Used in Indian Sign Language*". World Academy of Science, Engineering and Technology, ICIPCVPR 2015: 17th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Kuala Lumpur, Malaysia August 24 - 25, 2015 (Accepted)

## Papers in International Journals:

1. Daleesha M. Viswanathan, Sumam Mary Idicula. "*SVM Based Recognition of Facial Expressions Used In Indian Sign Language*." International Journal of Image Processing (IJIP) 9.1 (2015): 32.
2. Daleesha M Viswanathan, Sumam Mary Idicula,' *Recent Developments in Indian Sign Language Recognition: An Analysis'*, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 289-293
3. Daleesha M Viswanathan, Sumam Mary Idicula, '*Performance Evaluation of HOG for Recognition of Hand Gestures Alphabet with Resemblances'* International Journal of Information Processing, 9(2), 76-83, 2015,ISSN : 0973-8215

……….cs80……….

*Appendix* A

# ISL SIGNS

Teacher | Friend | Cat | Catch

Rat | Today | Read | Together

You | Food | Like | Ball

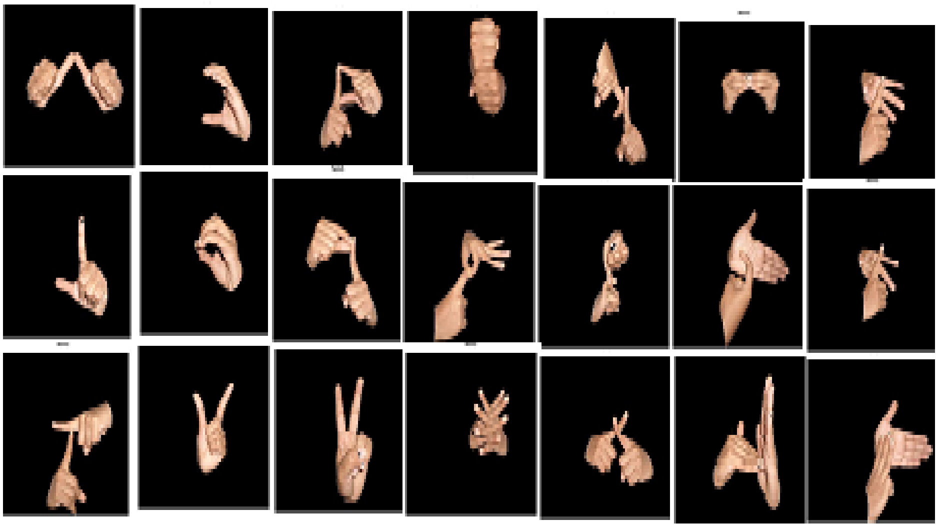
My | I | First | Part of Grandfather

Child Part of Sister/Brother Satureday Sunday

Monday Tuesday Wednesday Thursday Friday

Women Confusion Man Think

Black Blue Red



Double hand ISL alphabets.

……….. ……….